F=ma

E=mc$^2$

$$\int \frac{dy}{dx} dt$$

International Advanced Researches and Engineering Journal (IAREJ) is a double-blind peer-reviewed and publicly available online journal that has Editorial Board (https://dergipark.org.tr/en/pub/iarej/board). The editor in chief of IAREJ welcomes the submissions that cover theoretical and/or applied researches on **Engineering** and related science with Engineering. The publication language of the Journal is **English**. **Writing Rules** are given in Author Guidelines (https://dergipark.org.tr/en/pub/iarej/writing-rules). IAREJ publishes **original papers** that are research papers and technical review papers.

IAREJ publication, which is **open access**, is **free of charge**. There is no article submission and processing charges (APCs).

**IAREJ is indexed & abstracted in:**

Crossref (Doi prefix: 10.35860/iarej.)
Directory of Open Access Scholarly Researches (ROAD)
Directory of Research Journals Indexing (DRJI)
EBSCO
Google Scholar
Index Copernicus (ICI Journal Master List)
J-Gate
TUBITAK ULAKBIM TR Dizin (TR index)
WorldCAT

# TABLE of CONTENTS

## RESEARCH ARTICLES

*Research Article*

# Failure analysis of cast iron housing unit of UCF-216 type ball bearing

*Mustafa Guven Gok [a],\* iD and Mehmet Bunyamin Sahin [b] iD*

*[a]Department of Metallurgical and Materials Engineering, Gaziantep University, 27310, Gaziantep, Turkey*
*[b](Student) Department of Metallurgical and Materials Engineering, Gaziantep University, 27310, Gaziantep, Turkey*

| ARTICLE INFO | ABSTRACT |
|---|---|
| | One of the most common problems faced by briquette machine users in the industry is the failure of the cast iron housing unit of the UCF-216 bearings, to which the main shaft of the briquette machine is attached. In this study, the failure mechanism of a cast iron housing unit of a UCF-216 bearing broken during operation in a briquette machine was analyzed in order to develop a solution to the problem. Failure was occurring in the bolt hole areas of the housing unit. First of all, spectral analysis was performed and it was determined that the housing unit was grey cast iron. Then, the macrostructures of both unused and damaged housing units were examined. Some casting defects were detected in the bolt hole areas of the unused housing unit. It was also learned that during the assembly of the UCF-216 to the briquette machine, no torque meter was used while the bolts were tightened. In order to understand the effect of possible overtightening of the bolts, the system was modelled as a 3D solid. This model was exported to the finite element software and different bolt pretensions were applied considering the stresses created by the bolt tightening forces on the housing unit. As a result, it was understood that if the ratio of the stress caused by the axial bolt tightening force in the UCF-216 bearing unit to the yield stress of the UCF-216 housing unit is above 50%, fatigue failure will occur. |

## 1. Introduction

Concrete building block machines (CBBM) have been called by different names such as briquette machine, bims (pumice) machine, pumice machine and block making machine in the construction industry. This name difference is due to the purpose of using the CBBM. Because with the same machine, different products could be produced. By using several molds in the CBBM, construction materials with variable properties and dimensions such as interlocking briquettes, curbstones, hollow concrete blocks and pumice blocks can be produced. The most basic task of the CBBM is to shape the construction material in different molds using vibration and pressure. Actually, a CBBM is a kind of mass production machine. Because in this machine, which generally has an automation system, pumice, cement and water mixture in certain proportions is taken into molds and compressed with the help of vibration and hydraulic press [1–4].

The development of a CBBM with movable, adjustable dies that could make solid blocks dates back to the early 1900s. Harmon S. Palmer's "Machine for Molding Hollow Concrete Building Blocks" was patented in 1902 and made

a significant contribution to the development of the CBBMs known today [5,6]. Due to the fact that the construction materials produced in CBBMs have affordable prices, desired quality and repeatable features, these machines attract great attention on a global scale. Today, CBBMs are produced in both industrialized and developing countries. Türkiye is one of the important CBBM manufacturers and a significant part of the machines produced in Türkiye has been exported.

The CBBMs can be divided into 3 different categories as manual, semi-automatic and fully automatic. In addition, there are many types and models of CBBMs with different capacities. Molds can be produced and mounted according to the desired size and model for the product to be produced in the CBBM. The latest models of these machines usually consist of different basic parts, namely the main body, the picking robot, the belt separator, the elevator and other mechanisms. All of the CBBMs work with electrical energy, and semi and fully automatic machines are controlled by automation software. The upper rammer, which is an important part of the CBBM and located on the main body of the machine, allows the

arms to move the compression mold upwards. The arms are fixed to the main shaft on the main body of machine and this main shaft is connected to the machine body by means of UCF type bearings so that it can rotate around its own axis (see Figure 1). The UCF type bearings are fixed to the machine body with M22 or M23 bolts and nuts. Since there are 2 main shafts for the mold holding arms, 4 UCF type bearings are used in a machine [7]. However, it is reported that UCF type bearings are suddenly broken while working on the CBBM as a very frequent problem.
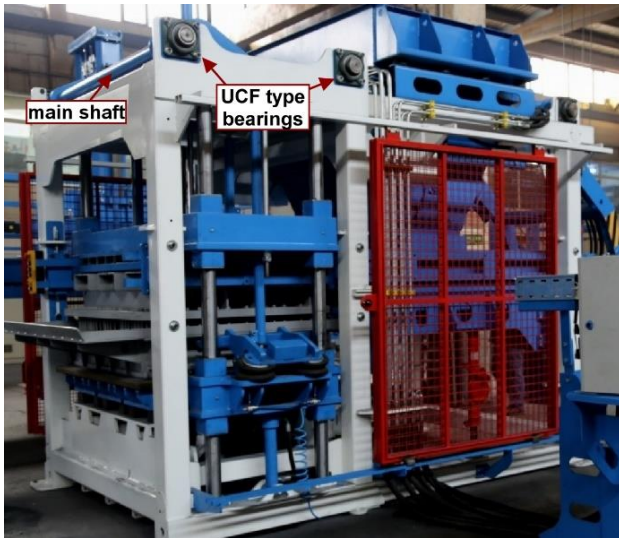


Figure 1. Main shaft and UCF type bearings on a fully automatic CBBM

The UCF series bearings are a square flange bearing type with four fixing holes and have a wide inner ring insert. Ball bearing units have international coding. The UCF coding on the housing unit used in this study indicates that the housing unit has a square flange type and cylindrical bore. The shaft diameter is coded in the numbers after UCF (UCF-216). Here the first digit is the diameter series of the bearing and the other digits are the size code of the bearing bore. They are one of the most preferred machine elements in the industry and have different codes according to their sizes. The UCF serial bearings consist of 2 main parts, a bearing and a cast iron housing. There is no doubt that the class and therefore the quality of the cast iron material that makes up the body (housing unit) of the UCF will vary according to the manufacturer [8–10]. In this study, the cast iron housing unit of the UCF-216 bearing, which was damaged during operation in the CBBM, was analyzed. Therefore, the aim of this study is to analyze the failure mechanism in the UCF-216 bearing housing unit, which is used in CBBMs and is damaged (broken) during operation, and to develop suggestions for solving the failure problem.

## 2. Material and Methods

### 2.1. Definition of failure

Basically, the problem was that the cast iron housing unit of the UCF-216 type ball bearings used in a fully automatic CBBM were broken. It was reported that the housing unit of the UCF-216 was damaged after the CBBM was used for a very short time. Moreover, the same problem was experienced both in the newly produced CBBM and after the replacement of the broken UCF-216 with a new one. The housing unit of the UCF-216 was made of cast iron. The UCF-216 housing unit was designed and produced in such a way that no damage would occur as a result of the load during operation under ideal conditions, and it had an infinite fatigue life. However, contrary to this situation, it was broken very often. The most important data we had was that the bolts were tightened without using a torque meter during the assembly of the UCF-216 to the CBBM. In order to understand and solve the problem, as explained in the next sections, both UCF-216 housing units were investigated macrostructurally and different bolt tightening scenarios were created in finite element software.



Figure 2. (a) Exploded 3D view of UCF and main shaft, (b) 2D technical drawing of the system.

Figure 3. (a) finite element model showing mesh elements, (b) loading and (c) symmetry conditions of the system.

### 2.2. Modelling of parts

For analysis, the 3D assembly diagram and 2D technical drawing of the system created according to the actual conditions and dimensions in the CBBM were given in Figure 2 (a) and (b), respectively.

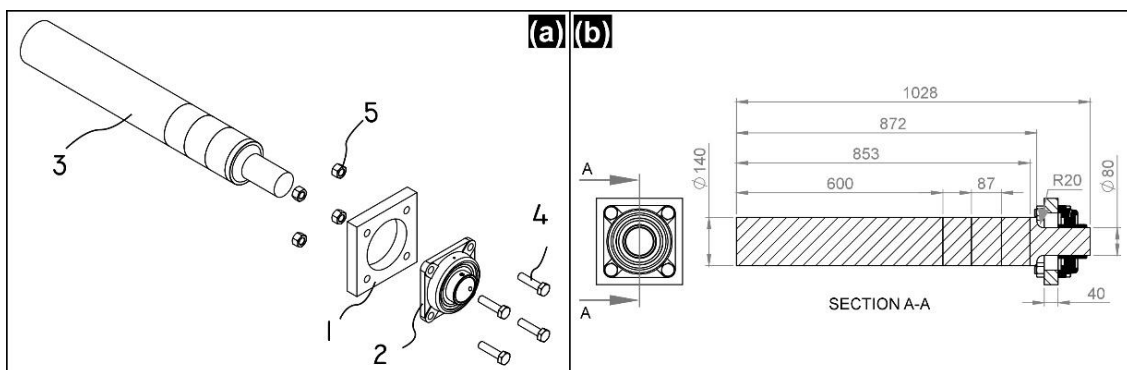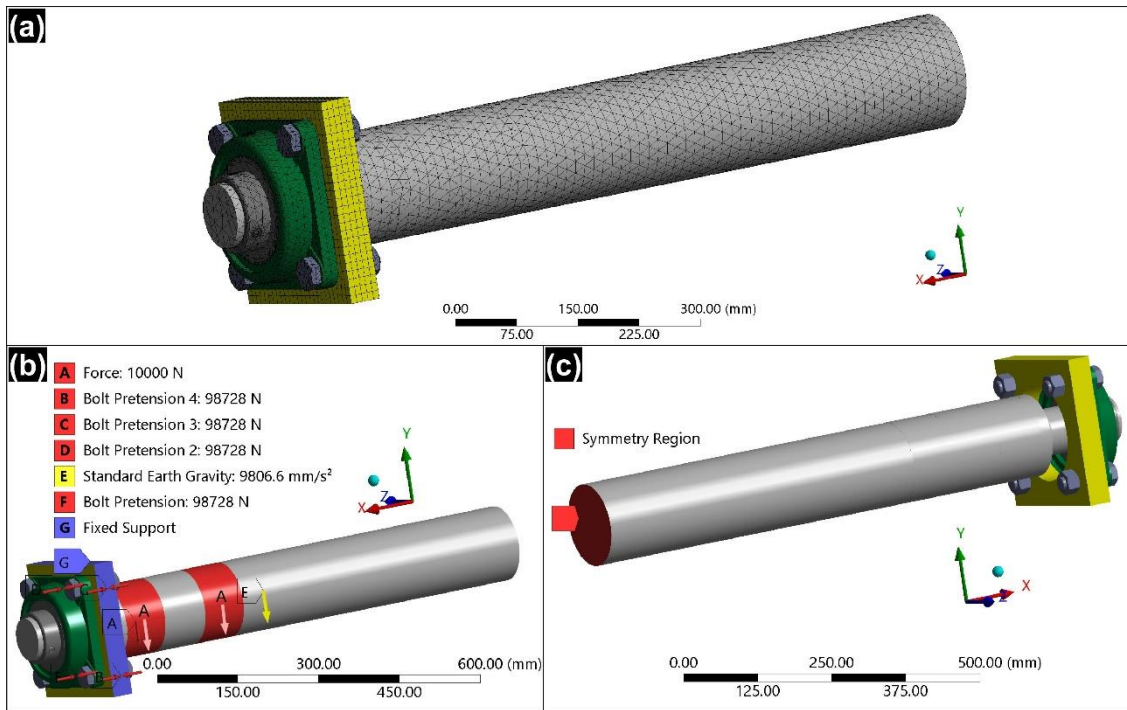SolidWorks software was used for the technical drawings and solid modelling of the parts of the system. The system consisted of five basic components. Parts 1, 2, 3, 4 and 5 shown in Figure 2 (a) were the headstock, UCF-216 bearing (with all its components), main arm, bolts and nuts, respectively. The UCF-216 (part number 2) consisted of a housing unit (made of cast iron) with 4 bolt holes for mounting and ball bearing system parts. Actually, the headstock (part number 1) was the body of the CBBM. In this part (part number 1), there were 4 bolt holes for mounting the UCF-216 (part number 2) and a hole in the middle to allows the main arm (part number 3) to pass contactless. The housing unit of UCF-216 (part number 2) was mounted to the headstock (part number 1) with 4 pieces M22 bolts (part number 4) and nuts (part number 5). Thanks to the bearing in the UCF-216, the main arm could rotate freely on its own axis. Thickness of the headstock was 40 mm. The diameter of the main arm mounted on the UCF-216 was 80 mm, while the diameter of the other part was 140 mm.

### 2.3. Spectral analysis of UCF-216

Spectral analysis was performed to determine the chemical composition of the UCF-216 housing unit. For this, an Oxford Instruments Foundry-Master Pro optical emission spectrometer was used. According to the spectral analysis results given as chemical composition in Table 1, the UCF-216 housing unit was made of grey cast iron. It is already stated in the catalogs of bearing manufacturing companies that UCF series bearing bodies are produced from gray cast iron material.

### 2.4. Numerical model, material, meshing, contact and boundary conditions

The 3D solid model of the system, which was created in accordance with its real dimensions and design using SolidWorks software, was exported to Ansys Workbench software. The material properties were defined as having linear elastic properties. On the other hand, it was very difficult or even impossible to define exact mechanical property values for cast irons. Because too many elements in the chemical composition of cast iron and the fact that they can be added to the composition in a wide range seriously affect its mechanical properties. In addition, especially the amount of graphite has a great effect on the stress-strain curve. While the tensile strength of grey cast irons was generally between 100 and 350 MPa, their compressive strength could be 3-4 times higher [11–13]. Therefore, in this study, consistent with studies in the literature [14–16], density, yield strength, ultimate strength, Young's modulus and Poission's ratio values for grey cast iron were chosen as 7.1 g/cm$^3$, 328 MPa, 480 MPa, 144.7 GPa and 0.28, respectively, since exact values must be defined in the software. As given in previous study [17], properties of AISI4140 steel were defined for bolts and nuts, and mechanical properties of structural steel were defined for other parts.

Table 1. Chemical composition (wt. %).

| Fe | C | Si | Mn | Cu | Ti | P | S | Bi | Cr | Ni | Mo | Others |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 93.0 | 3.77 | 2.26 | 0.541 | 0.116 | 0.096 | 0.091 | 0.077 | 0.016 | 0.01 | 0.005 | 0.002 | 0.016 |

After the material definition, the contact zones between neighboring components were determined. The contact surface of the UCF-216 housing unit with the headstock was defined as frictional (friction coefficient of 0.15 µ). As seen in Figure 3 (a), tetrahedron and hexahedral meshes were used to divide the system into finite elements. Different body sizing operations were applied to parts of the system in order to increase the mesh quality by approximating the mesh size distribution. As a result of the meshing process, the system contained a total of 174303 nodes and 107536 elements. In this study, the failure analysis of the housing unit of an UCF-216 bearing used in the CBBM was made. For this, as seen in Figure 3 (b), it has been determined that the maximum force acting on the UCF-216 by means of the main arm (part number 3) is 10000 N. The load was applied to the arm from two regions on the arm and in the -Y direction. These regions were the real places on the machine where the gripper arms were located and therefore the load was affected. On the other hand, the entire cross-sectional surface of the headstock was set as fixed support and a standard earth gravity in the -Y direction was applied to the entire system (see Figure 3 (b)). On the other hand, as the system is connected to the CBBM from the other side of the arm under the same conditions, as seen in Figure 3 (c), the cross-sectional surface of the midpoint of the arm length was defined as the symmetry region. Also, different bolt pretensions were applied to the bolts in order to see the effect of the bolt tightening force, since the UCF-216 was not assembled to the headstock with a torque meter. The stresses created by the axial bolt clamping force (*ABCF*) on the housing unit of UCF-216 ($\sigma_{ABCF}$) were taken into account when calculating the bolt pretensions.

For this, using Equation 1, bolt pretensions were applied to the bolts to create a stress equal to 0%, 10%, 30%, 50% and 70% of the yield strength of the housing unit of UCF-216 ($\sigma_{yUCF-216}$), and these were solved separately.

$$F_{cf} = \% \times \sigma_y \times \frac{\pi d^2}{4} \qquad (1)$$

Where *Fcf*, %, $\sigma_y$ and d were ABCF, the percentage of the desired stress value to be created on the UCF-216 housing unit, yield strength of UCF-216 housing unit and diameter of bolt, respectively. Therefore, the different ABCFs applied to the bolts were calculated by the ratio of the stress value created on the UCF-216 housing unit to the yield stress of UCF-216 housing unit ($\sigma_{ABCF}/\sigma_{yUCF-216}$), based on the area where the bolt head contacts the UCF-216 housing unit surface. Bolt pretension values applied to

the bolts were given in Table 2. Also as seen in Table 2, these ABCF values were converted to bolt tightening torque using Equation 2 [18,19]. These bolt tightening torque values were the values that should be used when tightening the bolts using a torque meter.

$$T = k \times F_{cf} \times d \times l \qquad (2)$$

Where T, k and l were tightening torque, constant that depends on the bolt material and size and lubrication factor (%), respectively. The assembly process of UCF-216 housing unit to the CBBM is usually done using uncoated bolts and without the use of lubricant. Therefore, in this study, the k value was taken as 0.2. In addition, the lubrication factor was neglected.

Table 2. Applied bolt pretension and bolt tightening torque values.

| $\sigma_{ABCF}/\sigma_{yUCF-216}$ (%) | Bolt pretension (N) | Bolt tightening torque (Nm) |
|---|---|---|
| 0 | 0 | - |
| 10 | 19746 | 87 |
| 30 | 59237 | 261 |
| 50 | 98728 | 435 |
| 70 | 138220 | 609 |

### 2.5. Analyzes

Macrostructural analyses were carried out to investigate the failure mechanism in the UCF-216 housing unit. For this, first of all, it was investigated whether there was a macrostructural defect during production on a randomly selected and unused UCF-216 bearing system. Then, the fractured surface of UCF-216 housing unit, which was damaged during working on the CBBM, was examined macrostructurally. In this study, deformation stress and fatigue analyses were carried out under given boundary conditions by using static structural module of Ansys finite element software. Since the aim of this study was to determine the failure mechanism in the UCF-216 bearing, the focus was on the housing unit of UCF-216 in the analysis of the deformation, stress and fatigue safety factor distributions. During the analyses, the ambient temperature was set to 22 °C. In addition to the 10000 N load acting on the system, the effect of different bolt pretensions was also analyzed. Especially, the maximum equivalent (von-Mises) stress and minimum fatigue safety factor values and their distributions in the UCF-216 housing unit were analyzed in detail. Fatigue analysis was performed according to Goodman stress theory.
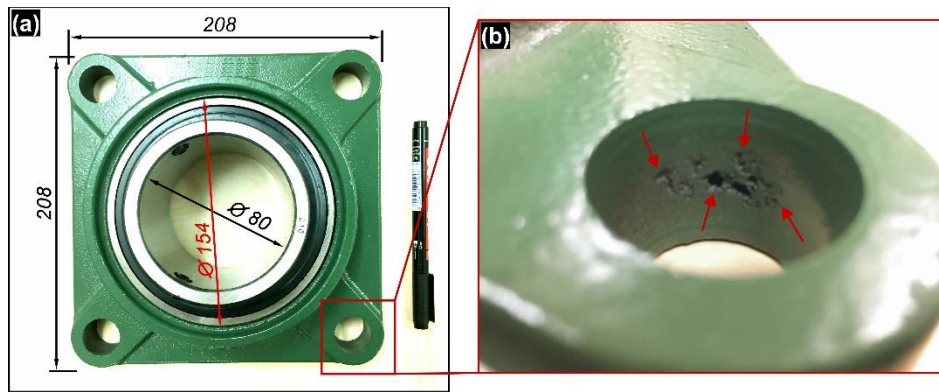
Figure 4. (a) photograph of a new and unused UCF-216, and (b) casting defects (porosities) on the inner surface of the bolt hole of this UCF-216.

## 3. Results and Discussion

### 3.1. Macrostructural analyses

Figure 4 (a) shows a macro photograph of an unused and randomly selected UCF-216 bearing. As can be seen from the photograph in the figure, any defects could be seen on the UCF-216 housing unit when viewed in general. However, when examined in detail, it was seen that there were voids having different diameters changing in between 4 and 1 mm on the inner surface of one of the bolt holes (indicated by arrows in Figure 4 (b)). Also, the voids had irregular shapes. These voids are thought to be casting cavities (defects) formed during the manufacture of UCF-216 housing unit. Dorula et al [20] produced and studied gray cast iron. The researchers observed irregularly shaped macroporosities in grey cast iron and determined that these were the result of shrinkage. The macroporosities observed by the researchers in grey cast iron were similar to the voids found in the bolt hole of UCF-216 housing unit in this study. Therefore, it was concluded that voids found on the on the inner surface of the bolt hole of UCF-216 housing unit occurred due to shrinkage during crystallization and solidification of cast iron. In this case, the use of appropriate inoculants has an important place in solving the problem. However, the voids were also likely to be gas porosity defects. The problem in this case could have been eliminated by using molding sand having high gas permeability or by removing excess moisture from the molding sand or by ensuring that the molding sand was not over rammed [21].

As a result, finding casting defects in mounting bolt holes of UCF-216 housing unit is a major problem; they have a notch effect, negatively affecting the mechanical properties and especially the fatigue strength. In Figure 5 (a), the macrostructure of UCF-216, which was damaged during operation in the CBBM, was given. The housing unit was broken in one of the bolt holes. In addition, when examined in general, it is seen that the damage did not cause a large number of broken pieces. The UCF-216 housing unit was broken in only two pieces. This could be

evidence that the damage was not caused by a force acting as an impact, but by a statically acting force. In addition, the fact that the failure occurred in the bolt hole region was proof of our theories of both casting defect and abnormal bolt tightening torque. Because, as indicated by the white arrows in Figure 5 (b), the damage probably caused by over-tightening the bolt was clearly noticeable. These damages were formed in the areas where the bolt contacted the housing unit and progressed into the bolt hole. Moreover, areas of porosity thought to be casting defects were also observed in Figure 5 (b) (shown in blue circle in Figure 5 (b)). Moreover, beach marks and macrocracks, which were thought to have occurred due to fatigue, were seen in Figure 5 (c). The rough and dimple-like morphology of these macrostructures was a sign that ductile fracture was dominant in the failure. Rihan [22] and Cullin [23] studied the fatigue failure of cast irons under stress and corrosion in different studies. Researchers observed fracture surface macrostructures having ductile fracture traces similar to those in this study.
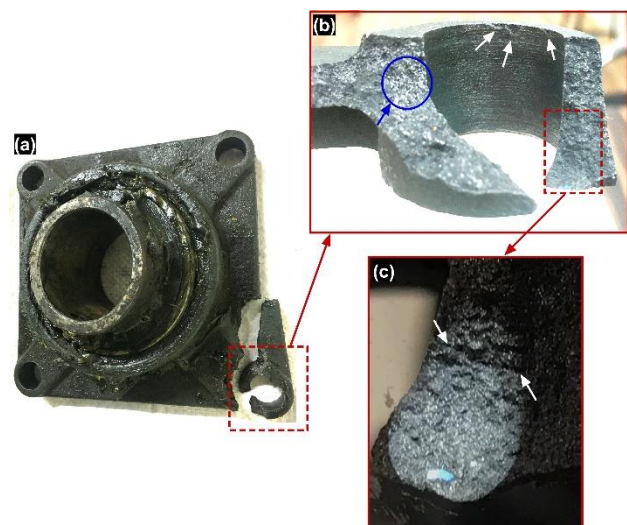


Figure 5. (a) Fracture in the housing unit of UCF-216, (b) general macrostructure of the broken piece, (c) traces of ductile fracture and fatigue failure.

### 3.2. Stress and fatigue analyses

As a result of different bolt pretensions applied to the bolts, the maximum von-Mises stress values on the UCF-216 ranged from 32.29 to 265.57 MPa. The graphical representation of these values were given in Figure 6.
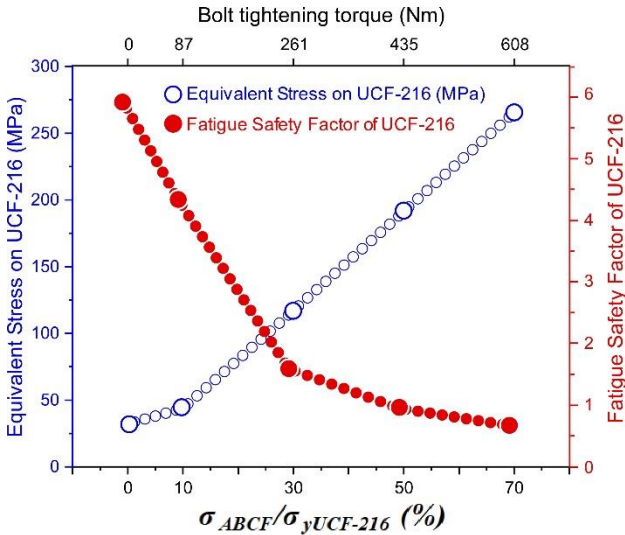


Figure 6. Graph of max. equivalent (von-Mises) stress and fatigue safety factor vs σABCF/σyUCF-216 (%) and bolt tightening torque.

Since all of these values were below the yield stress value of UCF-216, it was understood that the load applied to the system (10000 N) and the pretensions applied to the bolts (19746 – 138220 N) did not pose a problem on the UCF-216 in terms of static stress safety. When the factor of safety for stress (FOSS) coefficients obtained by dividing the yield stress of UCF-216 to the maximum von-Mises stress occurring on the UCF-216 under the specified conditions were calculated, it was understood that the lowest FOSS value was 1.24. However, this situation was valid for the materials that did not have manufacturing defects. As shown under heading 3.1 in this study, it will be almost impossible to calculate the FOSS values correctly in case of manufacturing defects such as casting voids in the UCF-216 body.

On the other hand, von-Mises stress distributions on the UCF-216 were given in Figure 7. Figure 7 (a) shows the stress distributions in the system without bolt pretension. As can be seen, the maximum stresses were concentrated on the bolts due to the shear effect. Because, as described in section 2.4, the contact surfaces were defined as frictional contact in the assembly process of UCF-216 to the headstock surrounded by fixed support. Furthermore, as seen in Figure 7 (b), the maximum stress on UCF-216 was 32.89 MPa without applying bolt pretensions to the bolts, and the stresses were concentrated in the bolt hole near the bolt head. When the bolts were tightened enough to create 10% of the yield stress of UCF-216 (19746 N) on the UCF-216, it was observed that the maximum stress on the UCF-216 was concentrated on the bolt hole interior surface (see Figure 7 (c)). As seen in Figure 7 (d-e), the regions where stresses were most intense in UCF-216 with the increase of applied bolt pretension were again the surfaces of the bolt hole near the bolt head. Therefore, the areas with the highest risk of fracture in UCF-216 were those close to the bolt holes. As explained in the previous section, in UCF-216, the region where the fracture occurred was the same as the regions where the maximum stresses were concentrated.
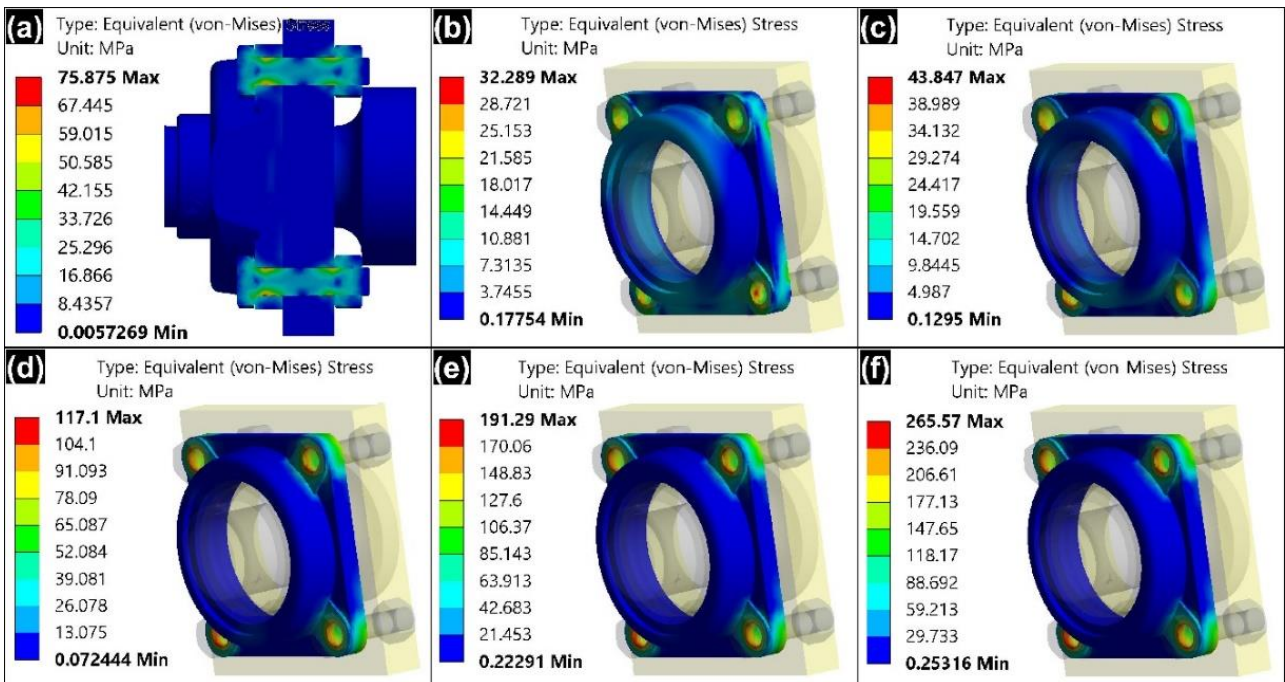


Figure 7. Distribution of equivalent (von-Mises) stress for different bolt pretensions (a) 0 N (for all components of system), (b) 0 N (for UCS-216), (c) 19746 N, (d) 59237 N, (e) 98728 N and (f) 138220 N.

Figure 8. Distribution of fatigue safety factor for different bolt pretensions; (a) 0 N, (b) 19746 N, (c) 59237 N, (d) 98728 N and (e) 138220 N.

However, as seen in Figure 8 (e), when the bolt pretension was increased to 138220 N, the minimum FOSF value decreased to 0.73. As it is known, if the FOSF value is below 1, the system does not have an infinite cycling life. In order to use the system saf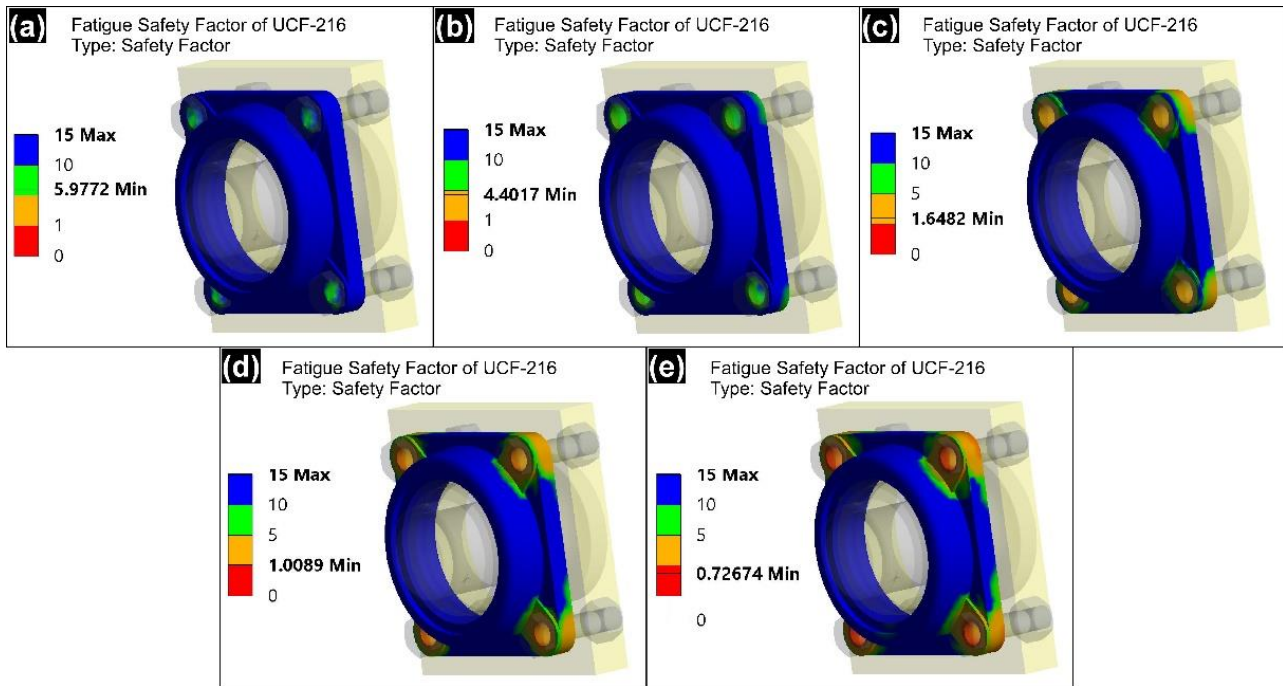ely, a maximum bolt pretension of 98728 N should be applied to the UCF-216 bolts (Figure 8 (d)). This meant that the stress created by the bolts in UCF-216 could be a maximum of 50% of the yield strength of UCF-216. In one study [24], researchers demonstrated that axial force from over-tightening causes deformation and fatigue damage in cylindrical roller bearings. At the same time, previous study [17,25] had also shown that excessive bolt pretension could cause premature fatigue damage in the material. On the other hand, the presence of manufacturing defects in the material will seriously and negatively affect the fatigue life of material. Stress will intensify in the regions where casting voids exist, and these voids will be the places where fatigue cracks start and propagate by showing notch effect in cyclic loading. Therefore, the risk of failure was significantly increased, when the UCF-216 had manufacturing defects such as casting voids, as in this study.

## 4. Conclusions

In this study, failure analysis of cast iron housing unit of an UCF-216 bearing used in a CBBM was performed. Unused and damaged UCF-216 housing units were macrostructurally examined and finite element modelling of their use in the CBBM was made by applying different bolt pretensions. Obtained results and recommendations were given in below.

- The cast iron housing of the unused UCF-216 contained casting defects. These casting defects were voids ranging from 4 to 1 mm in diameter and were on the inner surfaces of the bolt hole.

- Traces of damage caused by over-tightening were seen on the broken housing unit. In addition, both casting defects (voids) and beach marks of fatigue failure were observed in the macrostructure of the fractured surface.

- As a result of the finite element analysis, the highest von-Mises stress on the housing unit was found to be 265.57 MPa. Also, in all cases, stresses were concentrated in the bolt hole regions where failure occurred.

- It was concluded that the stress to be created on the UCF-216 housing unit by tightening the bolts can be up to 50% of the yield strength of the housing unit. The fatigue safety factor value became less than "1" when the bolts were tightened further. Minimum fatigue safety factor values were also concentrated in the bolt hole regions.

- Failure to the UCF-216 housing unit was caused by over-tightening of the bolts and casting defects also contributed.

- Appropriate inoculants could be used to eliminate casting defects from shrinkage. Moreover, casting defects caused by gas porosities could be eliminated by using molding sand having high gas permeability, by removing excess moisture from the molding sand or by ensuring that the molding sand is not compressed too much.

- The use of a torque-meter was essential when tightening the bolts

## Declaration

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

## Author Contributions

Mustafa Guven Gok: Writing, Finite Element Analyses, Visualization, Macrostructural Analyses, Review & Editing. Mehmet Bunyamin Sahin: Writing, Investigation of the industrial application where the UCF is used.

## References

1. Starovoytova, M.D., Namango, S.S., Arusei, D., *Innovative Conceptual Design of Manual-Concrete-Block- Making-Machine*. Innovative Systems Design and Engineering, 2016. **7**(7): p. 41–52.

2. Yemane Zemicheal., Qi Houjun., *Design, Analysis and Development of Improved Hollow Concrete Block Making Machine*. International Journal of Engineering Research and Technology, 2020. **9**(03): p. 298–302.

3. Onyeakpa, C., Onundi, L., *Improvement on the Design and Construction of Interlocking Blocks and its Moulding Machine*. IOSR Journal of Mechanical and Civil Engineering, 2014. **11**(2): p. 49–66.

4. Hall, J.P., *The Early Developmental History of Concrete Block in America*. Ball State University, Historic Preservation, Master Thesis. 2009: p. 1-73.

5. Palmer, H.S., *Machine for Molding Hollow Concrete Building Blocks*. United States Patent Office, 1903. Patent No: 727,427.

6. Charles, P., *Practical Concrete-Block Making*. 1908. USA: New York Industrial Publication Company.

7. E. Catalogue. [cited January 9, 2023]; Available from: https://www.esermakmakina.com/katalog.html.

8. Guyer, R.A., R*olling bearings handbook and troubleshooting guide*. 1996. USA: Radnor, Pennsylvania: Chilton Book Co.

9. Riduttori, B.S. p. A., *Gear Motor Handbook*. 1995. Berlin: Heidelberg: Springer.

10. Harris, T.A., Kotzalas, M.N., *Essential Concepts of Bearing Technology*. 2006. 6th ed., USA: Boca Raton, CRC Press.

11. Yakur, R., Çiftçi, Ö., *Investigation of the microstructure, hardness, and compressive properties of TaC-reinforced lamellar graphite cast irons*. European Mechanical Science, 2023. **7**(2): p. 56–62.

12. Yakut, R., *Investigation of Mechanical Properties of Grey Cast Irons Reinforced with Carbon Titanium Nitride (TiNC)*. Lubricants, 2023. **11**(454): p. 1–12.

13. Properties and characteristics of grey cast iron. [cited January 9, 2023]; Available from: https://www.zhycasting.com/properties-and-characteristics-of-grey-cast-iron/.

14. Ragul, G., Kalivarathan, G., Jayakumar, V., Maruthur, P., Jacob, I., Naveen Kumar, S., *An Analytical Investigation On Design And Structural Analysis Of Cam Shaft Using Solid Works And Ansys In Automobiles*. Indian Journal of Science and Technology, 2016. **9**(36): p. 1–9.

15. Kováčiková, P., Bezdedová, R., Vavro, J., Vavro, J., *Comparison of numerical analysis of stress-strain states of cast iron with vermicular graphite shape and globular graphite shape*. Procedia Engineering, 2016. **136**(): p. 28–32.

16. Chaturvedi, E., Dwivedi, R., *Design and Static Load Analysis Comparing Steel, Grey Cast Iron and Titanium Alloy as Materials for Breech Hinged Lugs in Recoil Weapons*. Advances in Military Technology, 2018. **13**(2): p. 265–75.

17. Gok, M.G., Cihan, O., *Investigation of Failure Mechanism of a DCI Engine Connecting Rod*. European Journal of Technique, 2021. **11**(2): p. 222–228.

18. Yu, Q., Zhou, H., Wang, L., *Finite element analysis of relationship between tightening torque and initial load of bolted connections*. Advances in Mechanical Engineering, 2015. **7**(5): p. 168781401558847.

19. Mínguez, J.M., Vogwell, J., *Effect of torque tightening on the fatigue strength of bolted joints*. Engineering Failure Analysis, 2006. **13**(8): p. 1410–1421.

20. Dorula, J., Kopyci, D., Guzik, E., Szczesny, A., *Procedure of Eliminating Porosity in Grey Cast Iron with Low Sulphur Content*. Materials, 2022. **15**(6273): p. 1–20.

21. Sertucha, J., Lacaze, J., *Casting Defects in Sand-Mold Cast Irons—An Illustrated Review with Emphasis on Spheroidal Graphite Cast Irons*. Metals, 2022. **12**(3): p. 1–80.

22. Rihan, R., Raman, R.K.S., Ibrahim, R.N., *Circumferential notched tensile (CNT) testing of cast iron for determination of threshold (KISCC) for caustic crack propagation*. Materials Science and Engineering A, 2005. **407**(1): 207–212.

23. Cullin, M.J., Petersen, T.H., Paris, A., *Corrosion Fatigue Failure of a Gray Cast Iron Water Main*. Journal of Pipeline Systems Engineering and Practice, 2015. **6**(2): p. 5014003.

24. Hou, X., Diao, Q., Liu, Y., Liu, C., Zhang, Z., Tao, C., *Failure Analysis of a Cylindrical Roller Bearing Caused by Excessive Tightening Axial Force*. Machines, 2022. **10**(5): p. 1–15.

25. Witek, L., Zelek, P., *Stress and failure analysis of the connecting rod of diesel engine*. Engineering Failure Analysis, 2019. **97**(): p. 374–82.

*Research Article*

# An optimization study focused on lattice structured custom arm casts for fractured bones inspiring additive manufacturing

*Mamoun Alshihabi [a],\* [iD] and Mevlüt Yunus Kayacan [a] [iD]*

[a]*Isparta Univeristy of Applied Sciences, Mechanical Engineering Department, Isprta 33220, Turkey*

| ARTICLE INFO | ABSTRACT |
|---|---|
| | In case of fractures, cracks or damage to bone tissues, it is important to use casts, fixatives and protective equipment. Especially in cases where long-term use of casts is required, soft tissue wounds may occur in the human body due to their moisture and airtight structure. For this reason, the use of casts with custom designs, breathable materials, and high mechanical properties has become widespread in recent years. This study focuses on the design of custom arm casts using advanced additive manufacturing technologies and lightweight materials. By utilizing Voronoi lattice structures and hexagonal surface meshes, optimized designs adaptable to additive manufacturing were obtained from a standard arm cast. All cast geometries were investigated under 196 N and 380 N forces. Then, the impact of a 100 g and 1000 g concrete piece with a speed of 12.5 m/s on the arm cast was investigated. As a result of the analyzes, stress, impact plate velocities, deformation, strain and deformation energy were evaluated. The results showed that the designed arm casts have up to 60% better impact strength compared to conventional arm casts. Based on the findings of this study, the use of custom arm casts with optimized lattice structures designed for additive manufacturing will demonstrate high performance. |

## 1. Introduction

Casting is a typical non-surgical procedure used to immobilize and stabilize shattered bones and promote normal healing [1]. It entails wrapping the wounded area in a stiff, hard substance to stabilize the bone, stop it from moving, and speed up the healing process [2]. Casts are often made of plaster or fiberglass and can be customized to fit the unique features of the injured limb or body part [3]. For many years, the application of casting to treat bone fractures has been a widely accepted and successful method. Casts not only assist in reducing pain and swelling, but also allow the broken bone to stay in a fixed position, promoting healthy bone formation. Accidents, falls, sports injuries, and underlying illnesses like osteoporosis are just a few of the causes of bone fractures [4]. Many fractures can be repaired without surgery utilizing techniques like casting, even though other fractures may necessitate surgical intervention [5]. Depending on the location and seriousness of the injury, different types of casts may be used for bone fractures. For instance, a short-arm cast, which covers the arm from the wrist to just below the elbow, may be necessary for a fractured arm [6]. On the other hand, a long-leg cast that

covers the upper thigh and toes may be necessary for a fractured leg [7]. Depending on the situation, a cast may be necessary for a few weeks in certain instances or for several months in others [8]. After the cast is taken off, physical therapy is frequently suggested to assist the injured area regain its strength, flexibility, and range of motion [9].

For years, people have used casts to immobilize and stabilize fractured bones. They were made from materials like mud, clay, and tree bark. Casts were later developed in the middle of the 19th century and swiftly took over as the preferred method for treating bone fractures [10]. Casts were created by encircling the wounded limb or body part with wet strips of cast of Paris, which would subsequently harden and offer support. Fiberglass castings were created in the middle of the 20th century as a lighter, more pleasant alternative to cast of Paris since they were heavy and may irritate the skin [11]. A fiberglass material is wrapped around the wounded area to create a fiberglass cast, which is subsequently hardened by the application of synthetic resin [12]. Modern casts are frequently made utilizing additive manufacturing technologies to get a precise and cozy fit that is custom-made to the patient's demands [13].

Furthermore, new materials like carbon fiber and thermoplastics have been created to make casts that are even lighter and stronger than conventional fiberglass casts. From the basic techniques of the past to the sophisticated, high-tech materials of the present, the development of casts used in bone fractures has come a long way [14]. These developments have improved casting as a therapy option for patients with bone fractures, making it safer, more efficient, and more comfortable.

Thermoplastic, fiberglass, and cast of Paris are common materials used to create modern arm casts. Because of its light weight, great strength, and stiffness, fiberglass is the material that is utilized the most frequently [15]. It can be shaped to the shape of the arm, giving a tight fit for the best immobilization. On the other hand, cast of Paris is more conventional and is occasionally still employed [16]. Although it takes longer to dry than fiberglass, it can offer a harder, more secure fit and it's a cheaper option than fiberglass. Thermoplastic castings are also growing in popularity since they are lightweight, waterproof, and can be heatedly molded for a customized fit [17]. Fiberglass casts are more resistant to breaking or cracking due to their higher tensile strength (305 MPa) and flexural modulus (17.2 GPa) in terms of mechanical qualities [18]. They are also less likely to bend or deform under load because to their higher flexural modulus. The cast of Paris, on the other hand, is better at absorbing shock and can distribute pressure more evenly across the arm while having a lower flexural modulus of 1.74 GPa and a much lower tensile strength of 2.76 MPa [19]. Thermoplastic casts have an impact strength of 130 J/m and an elastic modulus of 1.6 GPa, which is higher than plaster of Paris casts but lower than fiberglass. Because of their great impact strength, they are less prone to crack or break when dropped or bumped [20]. Additionally, they are more flexible and can more easily adapt to the curve of the arm due to their lower modulus of elasticity.

Arm cast production using additive manufacturing has been researched as an effective technique [21]. Compared to conventional casting methods, this technology has more design flexibility, customization, and shorter lead times [22]. A lightweight, custom-fitted cast that includes features like air holes and detachable components can be produced via 3D printing. Casts made using 3D printing have been demonstrated to be substantially lighter and more comfortable for patients while also having mechanical attributes that are comparable to those of conventional casts in terms of strength and rigidity [23]. Additionally, it has been discovered that using 3D printed casts with an open lattice structure improves airflow and lessens skin irritation [24]. Even if there are issues that need to be addressed, 3D printing for arm casts has promise. The price of 3D printing supplies and equipment, the length of time required for printing, and worries about

bacterial development in the porous lattice structures are a few of these drawbacks [25]. Despite these difficulties, 3D printed casts are a viable solution for the future of arm casting due to their prospective advantages [26].

Six distinct prosthetic arm cast models were created for this study, covering the region from the wrist to just below the elbow. These models include two porous cast designs, three designs reinforced with a Voronoi surface lattice on their outer layer, and one original body design that has not been optimized. Each design has an own unique set of Voronoi point count and unique thicknesses. The size and shape of the resulting Voronoi cells are strongly influenced by the Voronoi point count, which is the distance between the generating points. The resulting Voronoi cells will be small and asymmetric if the generating points are close to one another [27]. In contrast, large, uniform Voronoi cells will develop if the generating points are spaced widely apart [28]. Modern arm casts can be a successful treatment for a variety of arm fractures and injuries, but they also have their own problems [29]. The inability to customize the cast, which can lead to a poor fit and limited functionality, is one of the most frequent issues. Longer healing times and increased suffering may result from this [30]. Another frequent problem is skin irritation and rashes brought on by a lack of airflow and moisture build-up inside the cast [31]. The restrictive nature of the cast can also cause stiffness and limited mobility, which can weaken the muscles and cause poor circulation [32]. Due to the difficulty in reaching the skin underneath the cast, maintaining proper hygiene and washing it might be difficult as well [33]. In this study, we aim to address these common issues by designing casts that are strong, stiff, lightweight, easy to manufacture, and to improve hygiene two porous designs were incorporated.

This innovative study makes a significant contribution to the existing literature on orthopedic care by addressing the shortcomings of traditional arm casts through the application of advanced additive manufacturing technologies. By incorporating Voronoi lattice structures and hexagonal surface meshes, the research optimizes custom arm cast designs for enhanced breathability, adaptability to additive manufacturing processes, and superior mechanical properties.

## 2. Materials and Methods

Designing an arm cast that is optimized for maximum effectiveness was the main goal of this research. To accomplish this, five different arm casts with different designs and parameters were created. On a real-life model of a human limb, first the design for the arm cast was created. This first step involved closely analyzing the arm's anatomical characteristics, including the size and shape of the bones and muscles as shown in Figure 1. Using this knowledge, A cast was created to offer the best possible

support and security while facilitating healthy recovery.

A thorough optimization procedure was then applied to the resulting design. To obtain the best results, The design was iteratively modified and improved using innovative modeling techniques, resulting in the production of five different optimized versions of the arm cast using this method, each with a distinctive set of parameters and features as shown in Figure 2. A collection of optimized arm casts was created, which has the potential to enhance patient outcomes and quicken the healing process by fusing in-depth anatomical knowledge with advanced modeling methods. This was achieved through the ability to use advanced modeling methods and incorporate in-depth anatomical knowledge.
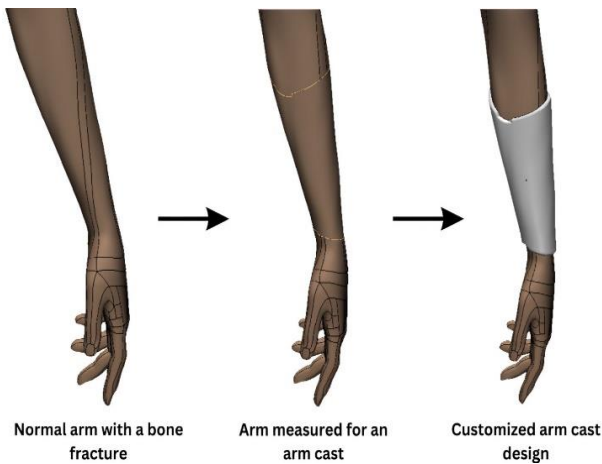


Figure 1. Process steps of the piece modeled from the broken arm
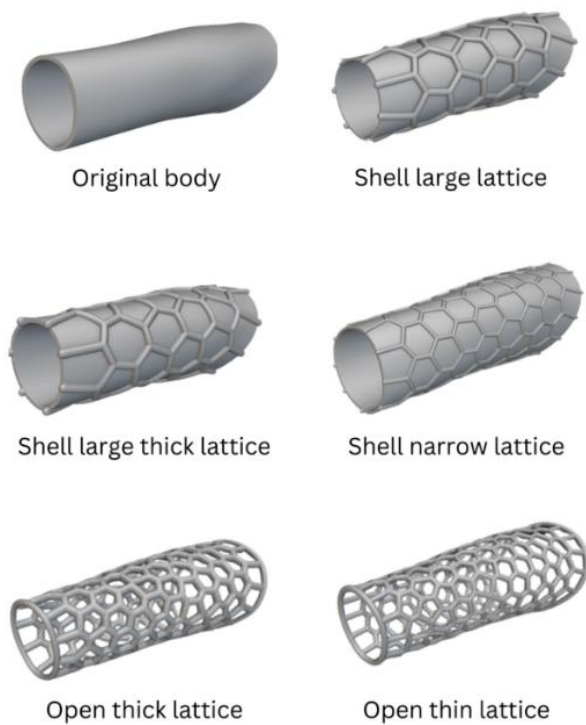


Figure 2. The optimized versions of the arm cast

With a thickness of 2 mm, the original arm cast design encompassed the region between the wrist and the elbow. Then, using nTopology, a program for 3D modeling and design optimization, this design was transformed into an implicit body. Implicit bodies, which are 3D geometries represented by a distance field, provide many advantages for design optimization [34]. Implicit bodies have several advantages over conventional CAD geometry, including greater flexibility and simplicity of manipulation. Additionally, they can be utilized to quickly and simply create a variety of design variations.

After the initial design was transformed into an implicit body, it was optimized to produce five distinct iterations of the arm cast. Three of these designs were based on the original body but had a Voronoi surface lattice reinforced on their outer surface. The shell large lattice 2 mm which is the first design comprised a lattice thickness of 6 mm, 50 Voronoi points, and 1 random seed. The second design that's named the shell large lattice 3 mm similar to the first, had a Voronoi point count of 50, a random seed of 1, and a lattice thickness of 6 mm. The original body's thickness, however, was increased to 3 mm. The third design which is the shell narrow lattice 2 mm had a lattice thickness of 3 mm, 100 Voronoi points, and one random seed. The fourth design which is the open large lattice 2 mm is a porous arm cast with a lattice thickness of 7 mm, 130 Voronoi points, and a random seed of 1. The fifth design which is the open narrow lattice 2 mm is a porous arm cast with a lattice thickness of 5 mm, 160 Voronoi points, and a random seed of 1. All of these parameters are shown in Figure 3 and listed in Table 1 where each design has been assigned a letter to be read easier.



Figure 3.
Lattice parameters
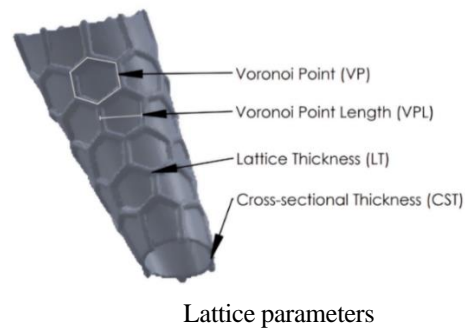
Table 1. nTopology design parameters

| Arm cast type | Voronoi point count | Voronoi point length | Lattice thickness | Cross-sectional thickness |
|---|---|---|---|---|
| (**A**) Original body | none | - | - | 2 mm |
| (**B**) Shell large lattice | 50 | 30 mm | 6 mm | 2 mm |
| (**C**) Shell large thick lattice | 50 | 30 mm | 6 mm | 3 mm |
| (**D**) Shell narrow lattice | 100 | 15 mm | 3 mm | 2 mm |
| (**E**) Open thick lattice | 130 | 12 mm | 7 mm | 2 mm |
| (**F**) Open thin lattice | 160 | 12 mm | 5 mm | 2 mm |

These improved designs provided a variety of solutions for various needs and preferences, with the goal of enhancing the arm cast's practicality and effectiveness. The two porous arm cast designs were made using the Voronoi surface lattice block. In contrast to the other ideas, this one included generating a surface lattice from the part's outside mesh without fusing the lattice to the original body. To ensure porosity, the lattice was maintained separately. To retain the arm cast's functionality, rims for the entrance and exit have to be designed. This was addressed by the addition of circular rims to both porous designs. These rims served as both a solid foundation for the lattice and a distinct boundary for the entrance and exit of the arm cast. After completing the designs they were exported from nTopology as mesh STL files each design was exported with a different mesh tolerance as they are not all similar. Some of them were exported with the mesh tolerance of 1.5 mm and some of them were exported with the tolerance 2.5 mm but the common factor between all of them is that all their meshes were simplified prior to exporting. The next step was to analyze the designs behaviours under different boundary conditions so for each design a static structural analysis and an explicit dynamics analysis was made. In the static structural analysis each design was fixed in a vertical line on the bottom of the arm cast and two forces were applied seperately on top of the cast in the exact opposite way of the fixed area as shown in Figure 4. The first force was 196 N and the second force was 380 N. after applying the forces and fixing the bottom the eqivalent Von-mises stress, the equivalent elastic strain and total deformation values were analyzed for each design.

To analyze the behavior of the cast designs after an impact, an explicit dynamics simulation was conducted on ANSYS Workbench. This involved dropping a concrete block onto the upper surface of the cast at a velocity of 12.5 m/s, using two different masses of 0.1 kg and 1 kg respectively with having a fixed support in a vertical line on the bottom face of the cast as shown in Figure 5. To carry out the simulation, all the cast designs were exported as meshes from nTopology and then imported into SolidWorks. then a rectangular extruded solid block was designed and positioned 0.5 mm above the upper surface of the arm cast, which was dropped onto the cast at the same velocity of 12.5 m/s.

The mathematical relationships used with ANSYS Workbench are presented within the scope of the study. In order to calculate explicit dynamics with Ansys, the following Eq. 1 was used.

$$y(t_{n+1}) = y(t_n) + \Delta t \dot{y}(t_n) \tag{1}$$

Eq. 2 calculates the directional strains along X axis. Note that directional strain is a measure of how much an object deforms or stretches along a specific axis.
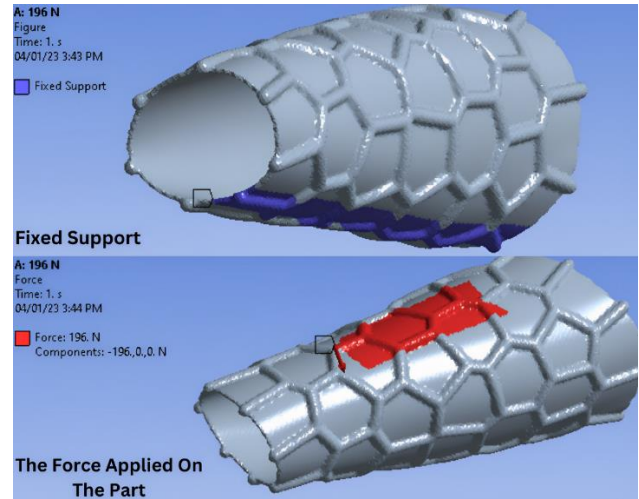


Figure 4. Boundary conditions of the static analysis
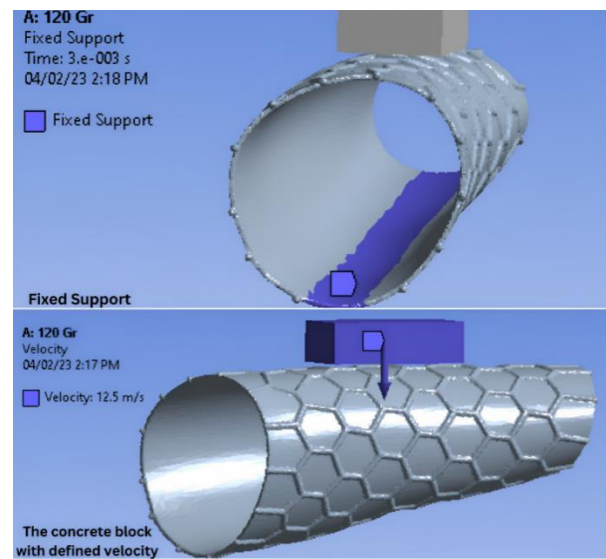


Figure 5. Boundary conditions of the explicit dynamics analysis

$$\varepsilon x = \left(\frac{\Delta Hx}{H}\right) \tag{2}$$

Eq. 3 is based on the principle of the work-energy theorem, which states that the work done on an object is equal to the change in its kinetic energy. In this case, the work done on the object is assumed to be due to a resistive force (such as friction) that causes the object to slow down from an initial velocity to a final velocity. The amount of kinetic energy lost by the object is proportional to the square of the velocity difference ($vi^2 - vf^2$) and the mass of the object (m).

$$\Delta KE = \left(\frac{1}{2}\right) * m * (vi^2 - vf^2) \tag{3}$$

Eq. 4 relates the reduction in kinetic energy of a material due to a directional strain in a particular direction so it's directional strain*kinetic energy reduction.

$$SK = \varepsilon * \Delta KE \qquad (4)$$

Eq. 5 relates the reduction in kinetic energy of a material due to a specific strain in a particular direction so it's specific strain*kinetic energy reduction.

$$SSK = \left(\frac{\Delta L}{L0}\right) * (\varepsilon * \Delta KE) \qquad (5)$$

According to the Courant-Friedrichs-Lewy (CFL) situation, the time steps needed depending on the unit mesh size were determined according to Eq. 6. "h" is characteristic length of a finite elements, "c" is wave speed in the material, "f" is safety factor as f≤1.

$$\Delta t \le f \left[\frac{h}{c}\right]_{min} \qquad (6)$$

Longitudinal Wave Speed value was obtained for materials according to Eq. 7. "E" is young's modulus and "ρ" is density. [35]

$$c = \sqrt{\frac{E}{\rho}} \qquad (7)$$

Basically, eigenvalue approaches were used in the calculations of the modal analysis according to Eq. 8. [M] is mass matrix and [K] is stiffness matrix. $[K] - \omega_i^2[M]$ is definition of natural frequencies and $\{\varphi\}_i$ is mode shapes.

$$[K] - \omega_i^2[M]\{\varphi\}_i = \{0\} \qquad (8)$$

Strain-rate hardening is represented by the equation, where K is the strain-rate hardening coefficient, and n is the strain-rate hardening exponent.

$$\sigma = K \varepsilon^{n+1} \qquad (9)$$

Damage evolution is described by the equation, where D is the damage variable, A is the damage coefficient, and m is the damage exponent.

$$D = 1 - \exp(-A \varepsilon^m) \qquad (10)$$

The stress-strain relationship is an equation, where σ is the stress in the material, E is the Young's modulus of elasticity, and ε is the strain in the material.

$$\sigma = E \varepsilon \qquad (11)$$

The contact forces equation, where $f_c$ is the contact force, $K_c$ is the contact stiffness, $\Delta u_n$ is the normal displacement, μ is the friction coefficient, $f_n$ is the normal force, and $\Delta u_t$ is the tangential displacement.

$$f_c = K c \Delta u_n + \mu f_n \Delta u_t \qquad (12)$$

The equation that represents the stress in the Johnson-Cook model is shown in Eq. 13. where $\sigma_F$ represents the flow stress of the material A describes the yield stress of the material under reference conditions B is the strain

hardening constant $\varepsilon_{pn}$ denotes the equivalent plastic strain $\varepsilon_p$ is the plastic strain rate $\varepsilon_0$ is the reference strain rate C is the strengthening coefficient of strain rate T represents the temperature m is the thermal softening coefficient of the model (in case of thermal effect consideration). [36]

$$\sigma F = (A + B \cdot \varepsilon p_n) \cdot \left(1 + \frac{\varepsilon_p}{\varepsilon_0}\right) C. (1 - T^m) \qquad (13)$$

The following equation represents the relationship between the flow stress and the defined strain rates and temperature-dependent flow curves. here $\sigma F$ is the flow stress $k_1$ is the Johnson-Cook material constant that describes the yield stress of the material under reference conditions $\varepsilon_p$ is the equivalent plastic strain $\dot{\varepsilon_p}$ is the equivalent plastic strain rate $k_t$ is the Johnson-Cook material constant that describes the strain hardening behavior of the material $T$ is the temperature $k_e$ is the Johnson-Cook material constant that describes the damage behavior of the material $D$ is the damage variable.

$$\sigma F = k^1(\varepsilon_p, \dot{\varepsilon_p}) \cdot k_t(\varepsilon_p, T) \cdot k_e(D) \qquad (14)$$

In the impact tests, the damage mechanisms of the materials were calculated according to the equation of state (EOS) principle and the following Eq. 15 was used.

$$P = K_1\mu + K_2\mu^2 + K_3\mu^3 \qquad (15)$$

The behavior of PLA (polylactic acid) under high strain rates and temperatures has been predicted using the Johnson-Cook model and PLA's characteristics. Thermoplastic known as PLA is biodegradable and renewable, and it is used in additive production among other sectors. Making arm casts out of PLA materials is a great application for additive manufacturing. The mechanical behavior of PLA under various loading scenarios, including tension, compressive, and shear loading, has been studied using the Johnson-Cook model which it's properties can be seen in Table 2. This is crucial when designing and enhancing the mechanical qualities of PLA-based products, such as arm casts, to satisfy the demands of the particular application.

Fused Deposition Modeling (FDM), Stereolithography (SLA), and Selective Laser Sintering (SLS) are a few techniques for 3D printing with PLA [38,39]. The most popular and reasonably priced PLA printing technique is FDM. The PLA filament used by FDM printers is melted and pushed through a tiny nozzle. The melted filament is then deposited by the printer in layers, building the item. The most economical way to print with PLA is through FDM because PLA is a reasonably inexpensive substance for 3D printing. The price of 3D printing with PLA, however, can vary significantly based on the size and complexity of the object as well as the caliber and features of the 3D printer being used.

Table 2. Properties of the material of the Johnson-Cook model [37].

| Property | Value | Unit |
|---|---|---|
| Density | 1240 | kg/m³ |
| **Isotropic Elasticity (PLA)** | **Value** | **Unit** |
| Young's Modulus | 4.1 | GPa |
| Poisson's Ratio | 0.35 | |
| Bulk Modulus | 6.05 | GPa |
| Shear Modulus | 1.58 | GPa |
| Yield Strength | 62.7 | MPa |
| **Johnson Cook Strength** | **Value** | **Unit** |
| Initial Yield Stress | 0.2 | GPa |
| Hardening Constant | 0.45 | GPa |
| Hardening Exponent | 0.2 | |
| Strain Rate Constant | 5 | |
| Reference Strain Rate | 1 | |

## 3. Results and Discussion

During the static analysis, the behavior of each design was thoroughly examined under two different load conditions. The first load was fixed at 196 N, and the second was 380 N. The focus of the study was on calculating the elastic strain, total deformation, and Von Mises stress values for each design. The equivalent stress values for each design under the 380 N force pressure are shown graphically in Figure 6. This information is essential for comprehending how each design reacts to outside loads and can guide choices regarding which design is most appropriate for particular uses. By examining the equivalent stress values, more information can be obtained regarding the possible strength and durability of each design.

Bar graphs were produced to demonstrate the results of the static analysis, showing the stress, strain, and deformation values for each design. These graphs give a thorough summary of the stress, strain, and deformation that each component underwent while operating within the previously mentioned boundary conditions. To emphasize the differences between the outcomes of each design, A method of comparison was used to represent each graph as a specific design. By comparing the graphs, the advantages and disadvantages of each design were determined, and more information was obtained about how they fared under the specified load circumstances.

First, a stress bar graph was constructed using data from each of the six parts. In order to make it simple to compare the various components, stress values from each design were chosen and plotted on a single graph. To show the findings of the static analysis under the two different load conditions, a bar graph was created. The graph as shown in Figure 7, depicts how each design behaves when subjected to a force of 196 N and 380 N.
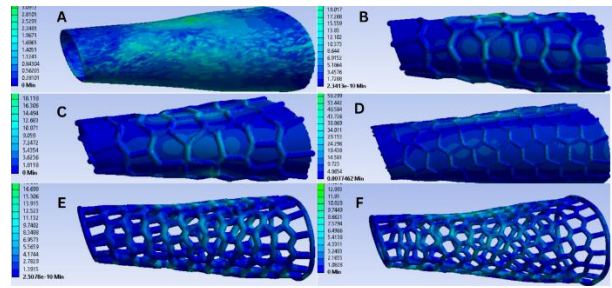


Figure 6. Von-Missses Stresses in MPa under 380 N of force
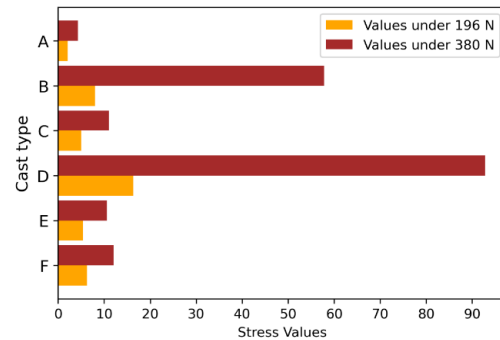


Figure 7. Maximum stress in MPa under 196 and 380 N of load

As seen in Figure 7 the open thin lattice (marked as F) showed a 2043.25% increase under 380 N of load which is a huge increase in stress values and the shell large lattice 2 mm (marked as B) showed a 1234.66% increase under 380 N of load which is also a very big increase in stress values which means these designs may not be durable compared to the other designs. In a similar study where the cast was subjected to only 30 N of load, the design experienced a maximum stress of 13.69 MPa. In contrast, the design D which showed the most stress and was subjected to a higher load of 196 N, experienced a maximum stress of approximately 18 MPa. However, when recalculating design D's stress as if it were subjected to the same 30 N load, the stress reduces to 2.75 MPa. When considering the difference in load conditions, designs in this study demonstrate superior stress resistance, highlighting their robustness in comparison to the aforementioned study [40]. Selecting strain data from each of the six parts, a strain bar graph was made. This graph has the maximum strain values from each design drawn on it to make it simple to compare the various components. It was made to display the results of the static analysis under the two different load circumstances. The maximum strain of each design is depicted in the graph in Figure 8 when the parts are subjected to a force of 196 N and 380 N.

In Figure 8 the shell narrow lattice (marked as D) showed 5262.6% increase in maximum strain values under 196 N of load and 3296.42% increase in maximum strain values under 380 N of load.
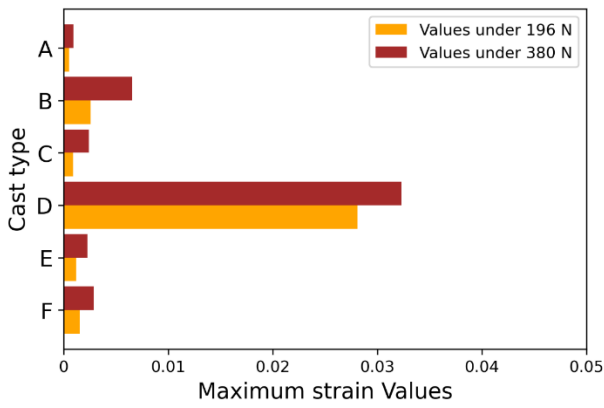
Figure 8. Maximum strain under 196 and 380 N of load
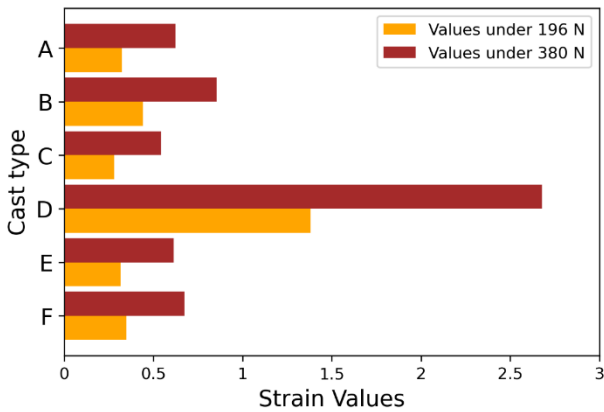


Figure 9. Deformation in mm under 196 and 380 N of load

A bar graph was created using information from all six designs to depict the overall deformation each component underwent during the static analysis. A visual representation that enables simple comparison between the various components by selecting the total deformation values from each design was created and plotted on a single graph. The results of the static analysis under the two different load circumstances were represented by the graph in shown in Figure 9 which shows strain values under a force of 196 N and 380 N.

In Figure 9 the shell narrow lattice (marked as D) yet again has proven to be not durable as it experienced the most deformation out of the other designs. It showed a 328.727% increase of deformation values under 196 N of load when compared to original body and It showed a 328.741% increase in deformation values under 380 N of load when compared to the original body. Based on the deformation distribution, the shell narrow lattice design exhibits significant deformation in some areas, indicating that this design is prone to deform under the applied load. This deformation renders the design unsuitable for use as an arm cast because it could pose a risk of harm to the patient's hand. It is important to note that the excessive deformation observed in the shell narrow lattice design is likely due to both its geometry and the specific load conditions applied during the analysis.
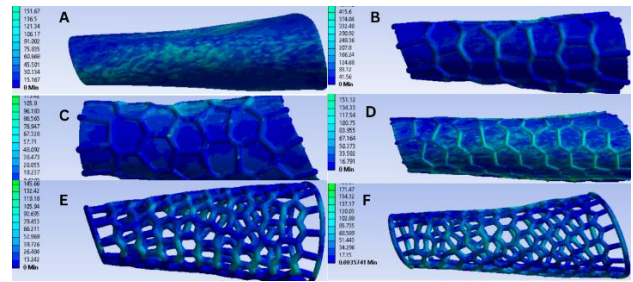


Figure 10. Von-Missses Stresses in MPa after impact with 1 kg concrete block
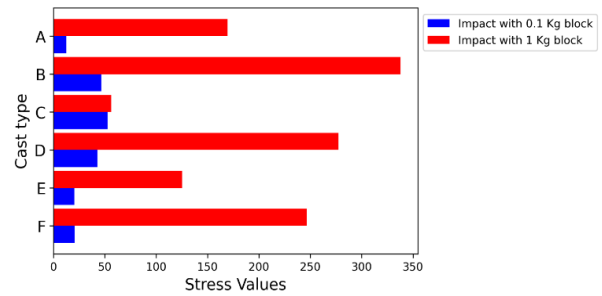


Figure 11. Von-mises stress in MPa for different impact scenarios on each part

Also, the open thick lattice and the shell large lattice 3 mm showed great results as the deformation values that were shown meant that these two cast types undergo very little deformation which protects the patient's arm from any impact. During the explicit dynamics analysis, The behavior of each design was investigated after dropping two different concrete blocks. The first block had a mass of 0.1 kg, and the second had a mass of 1 kg. During the analysis, the focus was on calculating the elastic strain, total deformation, and Von Mises stress values for each design. The equivalent stress values for each design under the 1 kg block are depicted in Figure 10. This information is crucial for understanding how each design responds to external loads and can aid in determining which design is best suited for specific applications. By examining the equivalent stress values, Insight into the potential strength and durability of each design can be gained through this analysis.

Bar graphs have been created to display the outcomes of the explicit dynamics analysis. These graphs make it simple to compare the numbers and results of the different parts and to understand how they behave. To show the stress values of each part in response to two distinct impact scenarios involving concrete blocks, a bar graph in particular has been made. The first the block weighs 0.1 kilograms, while the second block weighs 1 kilogram. This bar graph is shown in Figure 11.

As seen in Figure 11 the open thick lattice (marked as E) showed a 26.11% decrease in stress compared to the original body for the 1 kg impact analysis. And the shell large lattice 3 mm (marked as C) showed a 66.87% decrease in stress compared to the original body for the 1 kg impact analysis. The strain values of each component

under two different impact situations using concrete blocks have been displayed in an another bar graph. Figure 12's graph provides a clear and concise illustration of how each component responds to various levels of mass and impact and gives strain values to better understand the behavior of each part.

As seen in Figure 12 yet again the open thick lattice (marked as E) showed a 44.07% decrease in maximum strain when compare to the original body for 1 kg impact analysis. A 44.07% decrease in maximum strain value for an optimized part when compared to the original part indicates a significant improvement in the part's performance. In Figure 13 the total deformation values for all six parts are shown in a bar graph for both impact masses of 0.1 kg and 1 kg. The graph shows the strength of the open thick lattice (marked as E) and the shell large lattice 3 mm (marked as C) and how durable and strong these two parts are while on the other hand the shell narrow lattice yet again proved unsuitable as an arm cast because of the high level of deformation it shows. Compared to the original body.

All 5 optimized designs showed better deformation values than the original part but the open thick lattice (marked as E) and the shell large lattice 3 mm (marked as C) yet again showed a very significant reduction in deformation values for the 1 kg impact analysis as the open thick lattice showed a 53.96% decrease and the shell large lattice showed a 40.41% decrease. And in the 0.1 kg analysis the open thick lattice showed a 45.73% decrease. A velocity change line graph for each part for the impact with two different concrete blocks with the masses of 0.1 kg and 1 kg makes understanding the analysis better by showing how the velocity of a material changes over time due to an impact. Figure 14 and Figure 15 show the velocity changes of the concrete block for all the six designs.

When the plate velocities were analyzed, it was determined that open thick lattice (marked as E) and shell large lattice 3 mm lattice (marked as C) cast designs showed the best results. All lattice designs showed better results than the original body. As can be seen from the reduction in the velocities of the impacting plates, it is seen that the lattice structures have significant effects on impact damping. Impact energy calculations are crucial to impact analysis because they offer insightful knowledge into how materials behave under increased stress. The kinetic energy of the system is transferred during a collision, which causes the colliding objects to deform or sustain harm. The quantity of energy transferred during a collision and its impact on the objects are quantified with the aid of impact energy calculations. While directional strain refers to the direction of the deformation, specific strain refers to the amount of deformation that takes place in a particular region of the material.
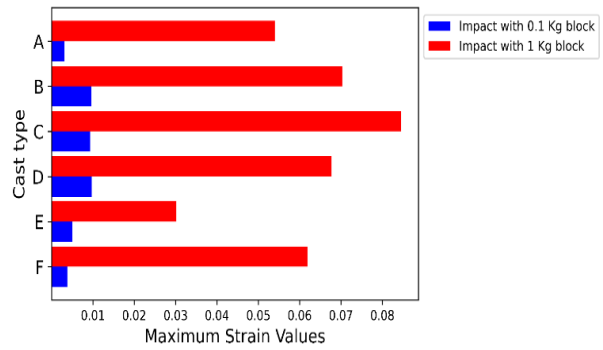


Figure 12. Strain for different impact scenarios on each
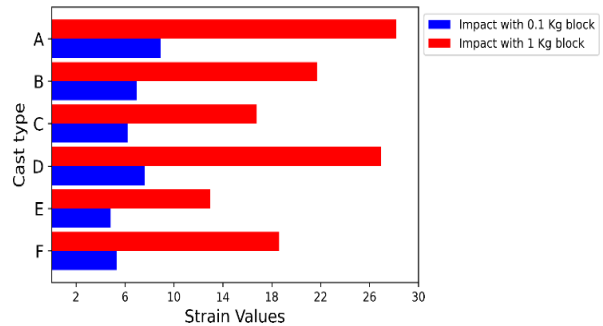


Figure 13. Deformation in mm for different impact scenarios on each part
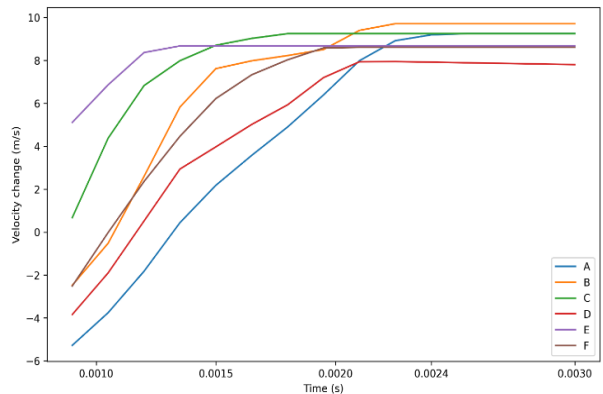


Figure 14. Velocity change line graph for all parts under 0.1 kg impact
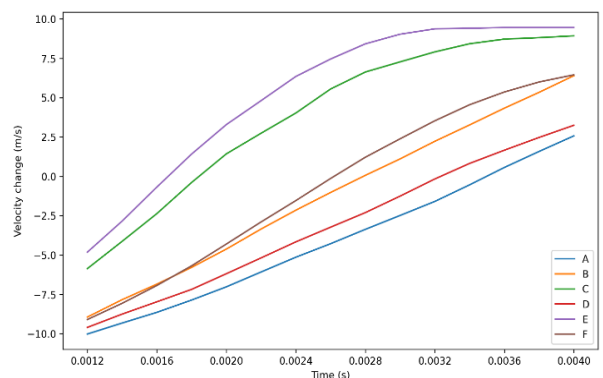


Figure 15. Velocity change line graph for all parts under 1 kg impact

In order to anticipate how materials will behave in upcoming impacts, it is essential to understand the specific and directional strain. Figure 16 shows the Kinetic energy reduction to directional strain values for each design for both 0.1 kg and 1 kg analyzes.

A graph of the kinetic energy against the specific strain was created to better comprehend how a material or structure behaves when it is deformed. A material experiences deformation or strain as a result of external pressures or loads, which lowers its kinetic energy. A graph depicting the relationship between the two variables was acquired by plotting the decrease in kinetic energy against the specific strain, which is the quantity of deformation per unit length or area. This graph can be used to build structures that won't fail under loads and deformations. The strength, stiffness, and ductility of the material properties that are crucial for creating effective and secure structures can be examined by analyzing the graph in Figure 17.
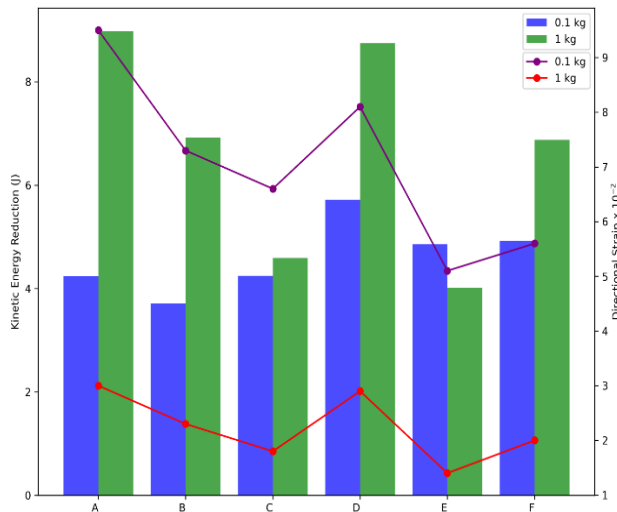


Figure 16. Kinetic energy reduction to directional strain
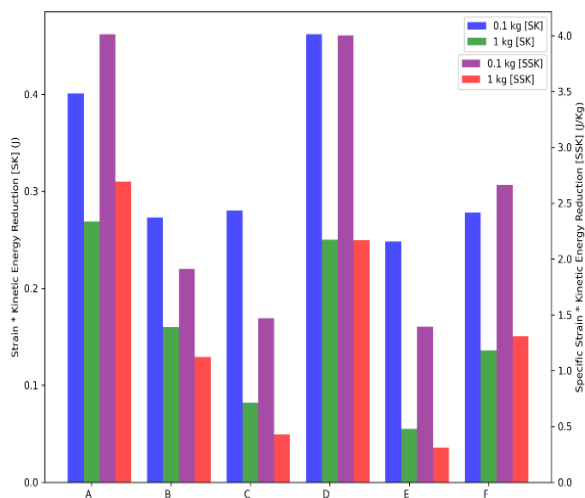


Figure 17. Kinetic energy reduction and energy analysis depending on specific strain

## 4. Conclusions

Within the scope of this study, the focus was on the optimization of casts of conventional designs with design parameters suitable for additive manufacturing. In this regard, hexagonal Voronoi lattice structures are designed volumetrically and superficially. All designs are analyzed under static loads of different strengths and under the impact of concrete blocks of different weights. According to the results of the analyzes, important findings are as follows:

- Under static loads, the original body is more advantageous in terms of stress and strain, whereas the lattice designs are advantageous in terms of deformations. It is important that a cast does not exert pressure on the damaged arm bone by deforming at the lowest level.
- In the case of impact of concrete blocks, open thick lattice and shell large lattice 3 mm casts gave good results in terms of stress, strain and deformations. It was understood that the lattice structures showed improved properties in impact moments.
- It was found that open thick lattice and shell large lattice 3 mm casts can reduce the impact velocities of concrete parts much faster. This may help to reduce the effects of the impact in a short time and prevent the pressure on the broken arm bone.
- Up to 85% lower SK and SSK data were obtained. It is understood that the rate of energy absorbed will increase with lower strain and kinetic energy drop. While the deformation change in gypsum is low, the rebound rate of the concrete block is also very high. In this way, the collision effect could be prevented effectively and quickly.

In future studies, customized cast design optimizations can be performed by using different lattice structures. Finite element analysis will be focused on the verification of the modelled casts by experimental methods.

## Declaration

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

## Author Contributions

In the scope of this study, Author 1 was the executor of the ideas, the optimizer of design, the creator of the analyses, concluder of results and writer of article. Author 2 was the creator of the idea, design boundary conditions and literature review and concluder of results.

## References

1. Kastenberger, T., Kaiser, P., Schmidle, G., Schwendinger, P., Gabl, M., and Arora, R., *Arthroscopic assisted treatment of distal radius fractures and concomitant injuries.* Archives of Orthopaedic and Trauma Surgery, 2020. **140**: p. 623-638.

2. Fang, Y., Yang, X., Lin, Y., Shi, J., Prominski, A., Clayton, C., and Tian, B. *Dissecting biological and synthetic soft–hard interfaces for tissue-like systems.* Chemical Reviews, 2021. **122**(5): p. 5233-5276.

3. Root, S. E., Sanchez, V., Tracz, J. A., Preston, D. J., Zvi, Y. S., Wang, K., and Whitesides, G. M., *An Expanding Foam-Fabric Orthopedic Cast.* Advanced Materials Technologies, 2022. **7**(9): p. 2101563.

4. Craig, J., Clarke, S., and Moore, P, *Orthopaedic and Trauma Nursing: An Evidence-based Approach to Musculoskeletal Care.* Principles of Fracture Management, 2023. p. 240-255.

5. Leixnering, M., Rosenauer, R., Pezzei, C., Jurkowitsch, J., Beer, T., Keuchel, T., and Quadlbauer, S, *Indications, surgical approach, reduction, and stabilization techniques of distal radius fractures.* Archives of Orthopaedic and Trauma Surgery, 2020. **140**: p. 611-621.

6. Dib, G., Maluta, T., Cengarle, M., Bernasconi, A., Marconato, G., Corain, M., and Magnan, B, *Short arm cast is as effective as long arm cast in maintaining distal radius fracture reduction: Results of the SLA-VER noninferiority trial.* World Journal of Orthopedics, 2022. **13**(9): p. 802.

7. Sabeh, K., Aiyer, A., Summers, S., and Hennrikus, W., *Cast application techniques for common pediatric injuries: A review.* Current Orthopaedic Practice, 2020. **31**(3): p. 277-287.

8. Farrell, S., Schaeffer, E. K., and Mulpuri, K., *Recommendations for the care of pediatric orthopaedic patients during the COVID pandemic.* The Journal of the American Academy of Orthopaedic Surgeons, 2020. **28**(11): p. 477–486.

9. Byrchak, V., Duma, Z., and Aravitska, M., *Effectiveness of the active physical therapy in restoring wrist and hand functional ability in patients with immobility-induced contracture of the wrist joint complicated by median nerve entrapment owing to distal forearm fracture.* Journal of Physical Education and Sport, 2020. **20**(6): p. 3599-3606.

10. Marin, E., Boschetto, F., and Pezzotti, G., *Biomaterials and biocompatibility: An historical overview.* Journal of Biomedical Materials Research Part A, 2020. **108**(8): p. 1617-1633.

11. Rezaei, R., *The Easy Wrap Orthopedic Cast*, Rochester Institute of Technology, 2017.

12. Varivodov, V. N., Kovalev, D. I., Zhulikov, S. S., Golubev, D. V., Romanov, V. A., and Mirzabekyan, G. Z., *Technological aspects of the use of cast polymer insulation for high-voltage switchgear and busbars.* Power Technology and Engineering, 2021. **54**: p. 915-922.

13. Paterson, A. M., Bibb, R., Campbell, R. I., and Bingham, G., *Comparing additive manufacturing technologies for customised wrist splints.* Rapid Prototyping Journal, 2015. **21**(3): p. 230-243.

14. Henkel, J., Woodruff, M. A., Epari, D. R., Steck, R., Glatt, V., Dickinson, I. C., and Hutmacher, D. W., *Bone regeneration based on tissue engineering conceptions—a 21st century perspective.* Bone research, 2013. **1**(1): p. 216-248.

15. Iftekhar, A., *Standard handbook of biomedical engineering and design.* Biomedical composites, 2004. p. 1-17.

16. Edwards, M. A, *Guide to Modelling in Clay and Wax: And for Terra Cotta, Bronze and Silver Chasing and Embossing, Carving in Marble and Alabaster, Moulding and Casting in Cast-Of-Paris or Sculptural Art Made Easy for Beginners.* Read Books Ltd, 2016.

17. Maji, P., and Naskar, K., *Styrenic block copolymer-based thermoplastic elastomers in smart applications: Advances in synthesis, microstructure, and structure–property relationships—A review*, Journal of Applied Polymer Science, 2022. **139**(39): p. 52942.

18. Gogoi, R., Niyogi, U. K., Alam, M. S., and Mehra, D. S, *Study of effect of NCO/OH molar ratio and molecular weight of polyol on the physico-mechanical properties of polyurethane cast.* World Applied Sciences Journal, 2013. **21**(2): p. 276-283.

19. Parmar, A. J., Tyagi, S. K., Dabas, V. S., Mistry, J. N., Jhala, S. K., Suthar, D. N., and Bhatti, I. M., *Assessment of the physical and mechanical properties of cast of Paris bandage cast used as a splinting and casting materials.* Veterinary World, 2014. **7**(12): p. 1123-1126.

20. Gibson, I., Rosen, D., Stucker, B., Khorasani, M., Gibson, I., Rosen, D., and Khorasani, M., *Materials for additive manufacturing.* Additive Manufacturing Technologies, 2021. p. 379-428.

21. Hasanov, S., Alkunte, S., Rajeshirke, M., Gupta, A., Huseynov, O., Fidan, I., and Rennie, A., *Review on additive manufacturing of multi-material parts: progress and challenges.* Journal of Manufacturing and Materials Processing, 2021. **6**(1): p. 4.

22. Pal, A. K., Mohanty, A. K., and Misra, M., *Additive manufacturing technology of polymeric materials for customized products: recent developments and future prospective.* RSC Advances, 2021. **11**(58): p. 36398-36438.

23. Banga, H. K., Kumar, R., Channi, H. K., and Kaur, S., *Parametric design and stress analysis of 3D printed prosthetic finger.* In Innovative Processes and Materials in Additive Manufacturing, 2023. p. 57-80.

24. Zolfagharian, A., Gregory, T. M., Bodaghi, M., Gharaie, S., and Fay, P., *Patient-specific 3D-printed splint for mallet finger injury.* International Journal of Bioprinting, 2020. **6**(2): p. 259.

25. Peng, S., Guo, Q., Thirunavukkarasu, N., Zheng, Y., Wang, Z., Zheng, L., and Weng, Z., *Tailoring of photocurable ionogel toward high resilience and low hysteresis 3D printed versatile porous flexible sensor.* Chemical Engineering Journal, 2022. **439**: p. 135593.

26. Richa, S., Bhaskar, J., and Kumar, A., *A Review on: 3D Printed Orthopaedic Cast for Improved Forearm Fracture*

*Rehabilitation*. International Journal for Research in Applied Science & Engineering Technology, 2021. **9**(11): p. 66-72.

27. Lei, H. Y., Li, J. R., Xu, Z. J., and Wang, Q. H., *Parametric design of Voronoi-based lattice porous structures*. Materials and Design, 2020. **191**: p. 108607.

28. Bukenberger, D. R., Buchin, K., and Botsch, M., *Constructing L∞ Voronoi Diagrams in 2D and 3D.* In Computer Graphics Forum, 2022. **41**(5): p. 135-147.

29. Shirley, E. D., Maguire, K. J., Mantica, A. L., and Kruse, R. W., *Alternatives to traditional cast immobilization in pediatric patients*. JAAOS-Journal of the American Academy of Orthopaedic Surgeons, 2020. **28**(1): p. 20-27.

30. Graham, J., Wang, M., Frizzell, K., Watkins, C., Beredjiklian, P., and Rivlin, M., *Conventional vs 3-dimensional printed cast wear comfort*. Hand, 2020. **15**(3): p. 388-392.

31. Williams, R. J., *Exploring thermal discomfort amongst lower-limb prosthesis wearers.* UCL University College London, 2020.

32. Craig, J., Clarke, S., and Moore, P., *Orthopaedic and Trauma Nursing: An Evidence-based Approach to Musculoskeletal Care*. Principles of Fracture Management, 2023.  p. 240-255.

33. Gao, Y., Yu, L., Yeo, J. C., and Lim, C. T*., Flexible hybrid sensors for health monitoring: materials and mechanisms to render wearability*. Advanced Materials, 2020. **32**(15): p. 1902133.

34. Patel, A. H., Baxi, N. J., and Gurrala, P. K., *A study on triply periodic minimal surfaces: A case study*. Materials Today: Proceedings, 2022. **62**: p. 7334-7340.

35. Satkar, A. R., Mache, A., and Kulkarni, A., *Numerical investigation on perforation resistance of glass-carbon/epoxy hybrid composite laminate under ballistic impact*. Materials Today: Proceedings, 2022. **59**: p. 734-741.

36. Aslam, M. A., Ke, Z., Rayhan, S. B., Faizan, M., and Bello, I. M., *An investigation of soft impacts on selected aerospace grade alloys based on Johnson-Cook Material Model.* In Journal of Physics: Conference Series, 2020. **1707**(1): p. 012008.

37. Pinto, V.C., Ramos, T., Alves, S., Xavier, J., Tavares, P., Moreira, P.M.G.P. and Guedes, R.M., *Comparative failure analysis of PLA, PLA/GNP and PLA/CNT-COOH biodegradable nanocomposites thin films*. Procedia Engineering, 2015. **114**: p.635-642.

38. Tanabi, H., *Investigation of the temperature effect on the mechanical properties of 3D printed composites*. International Advanced Researches and Engineering Journal, 2021. **5**(2): p. 188-93.

39. Bolat, ç., And ergene, b., *An experimental effort on impact properties of polylactic acid samples manufactured by additive manufacturing*. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 2023. **11**(2): p. 998-1013.

40. Mian, S.H., Umer, U., Moiduddin, K. and Alkhalefah, H., *Finite Element Analysis of Upper Limb Splint Designs and Materials for 3D Printing*. Polymers, 2023. **15**(14): p. 2993.

*Research Article*

# Detection of wheeze sounds in respiratory disorders: a deep learning approach

Leen Hakki [a] and Gorkem Serbes [b,*]

ᵃDepartment of Biomedical Engineering, Yildiz Technical University, Istanbul, Turkey
ᵇDepartment of Biomedical Engineering, Yildiz Technical University, Istanbul, Turkey

ABSTRACT

Respiratory disorders, including chronic obstructive pulmonary disease (COPD) and asthma, are major causes of death globally. Early diagnosis of these conditions is essential for effective treatment. Auscultation of the lungs is the traditional diagnostic method, which has drawbacks such as subjectivity and susceptibility to environmental interference. To overcome these limitations, this study presents a novel approach for wheeze detection using deep learning methods. This approach includes the usage of artificial data created by employing the open ICBHI dataset with the aim of improving in generalization of learning models. Spectrograms that were obtained as the output of the Short-Time Fourier Transform analysis were employed in feature extraction. Two labeling approaches were used for model comparison. The first approach involved labeling after wheezing occurred, and the second approach assigned labels directly to the time steps where wheezing patterns were seen. Wheeze event detection was performed by constructing four RNN-based models (CNN-LSTM, CNN-GRU, CNN-BiLSTM, and CNN-BiGRU). It was observed that labeling wheeze events directly resulted in more precise detection, with exceptional performance exhibited by the CNN-BiLSTM model. This approach demonstrates the potential for improving respiratory disorders diagnosis and hence leading to improved patient care.

## 1. Introduction

Among the leading causes of death worldwide are respiratory diseases, including chronic obstructive pulmonary disease (COPD), and asthma. COPD, the third leading cause of death globally, was responsible for 3.23 million deaths in 2019. In the same year, approximately 262 million people were affected by asthma, resulting in the deaths of 455,000 individuals [1-3]. The given statistics highlight the importance of early detection of respiratory diseases. The most frequently used method for respiratory disorders diagnosis is lung auscultation, which is a non-invasive and cost-effective method of diagnosis used to assess the condition of the lungs' health [4]. The auscultation method is used to listen to the sounds produced by the lungs. This method is essential for evaluating patients' respiratory symptoms (i.e. coughing, wheezing, or crackling). Although medical technologies for pulmonary diagnosis (i.e. spirometry, and chest X-ray) have advanced significantly, auscultation remains one of the commonly employed approaches for diagnosing respiratory sounds employing the traditional analog stethoscope [5]. Although auscultation with a stethoscope

is valuable, it has some limitations, such as its subjectivity, as it relies on the expertise of the physician. Additionally, it provides inaccurate information when used in noisy environments. Besides, there is a risk of infection if it comes into direct contact with the patient [4]. Due to these limitations, researchers sought to improve the efficiency of auscultation by parameterizing lung sounds through computerized lung sound analysis or digital stethoscopes, which include sampling, filtering, feature identification, and lung sound classification [6].

Respiratory sounds are produced by the air moving through the lungs and airways during breathing [7]. The respiratory sounds can be categorized into two classes; normal and abnormal (adventitious) sounds. Normal respiratory sounds are characterized by a low noise in the inspiration phase and are difficult to hear during the exhalation phase. Their highest frequency range is under 100 Hz [7]. On the other hand, extra respiratory sounds that are not normally heard during breathing are called adventitious sounds. Those abnormal (adventitious sounds) can indicate the presence of pulmonary ailment [8]. The Internal Lung Sound Association has divided

adventitious sounds into continuous and discontinuous sounds. The adventitious sounds are further classified into wheeze and rhonchi, which are continuous sounds, and fine and coarse crackles, which are discontinuous sounds [9]. Crackles are discontinuous and explosive clicking or crackling sounds caused by the opening of small airways. The duration of crackles is short and usually less than 100 ms [9]. Crackles can indicate various health conditions such as pneumonia, chronic bronchitis, bronchiectasis, congestive heart failure, and obstructive pulmonary disease [10]. Wheezes are continuous sounds that occur as a result of air passing through narrow passageways due to blockage in the airways [9]. The duration of wheezes is considerably longer than the duration of crackles. Their duration lasts more than 100 ms, with an average of 250 ms, and they have a dominant frequency of 100 Hz or greater [9]. The sound of wheezes can differ among individuals and is influenced by factors such as the extent of the condition and the location where the stethoscope was positioned during auscultation [11]. Wheezes can indicate various health conditions such as asthma and bronchial stenosis [7]. A wheezing sound can be classified as monophonic or polyphonic. Monophonic wheeze refers to the wheezing sound heard consistently throughout the respiratory cycle with a uniform pitch. It is generally caused by airway narrowing because of a foreign body or tumor. Polyphonic wheezing involves multiple wheezes of different pitches occurring simultaneously, which is generally caused by asthma or COPD [7]. The presence of such different wheeze types makes wheeze detection more complex.

The human's ear ability to indicate adventitious respiratory sounds is limited. The intensity and type of abnormal sounds, as well as their amplitude, are significant factors contributing to detection errors in respiratory signal analysis. In addition, other environmental factors such as artifacts coming from patient movements, coughing, or speech are also limiting factors for lung auscultation. Hence, the validation of automatic adventitious sound detection algorithms should not solely rely on auscultation as the standard diagnosis method. Based on the reasons discussed, there is a need for the development of methods and algorithms that can accurately detect the adventitious sounds and implement them into smart stethoscopes for more effective diagnosis of respiratory illnesses. Therefore, in the last years, researchers put their efforts into developing classification models using machine learning-based algorithms. Before 2017, the available data on respiratory sounds were limited and insufficient [12]. Later, a challenge for lung sound classification was published at the International Conference on Biomedical and Health Informatics (ICBHI) 2017, containing 920 respiratory records collected from 126 subjects [13].

In this study, artificial data derived from the ICBHI dataset was generated to address issues like the scarcity of wheeze events and excessive noise in certain respiratory recordings. The spectrograms obtained from the short-time Fourier Transform (STFT) analysis of the recordings were utilized for feature extraction. Prior to inputting these spectrograms into the constructed model, two labeling approaches were employed. While prior studies mostly used convolutional neural networks (CNNs) for lung sound classification, this study employs a convolutional recurrent neural network (CRNN) architecture to better capture time-dependent patterns in the data. Detecting wheeze sounds not only indicates their presence but also their timing during breathing cycles. This research seeks to develop an efficient algorithm using digital auscultation recordings for improved respiratory disease detection with smart stethoscopes. While previous studies focused on the classification of respiratory sounds as will be discussed in the next section, this study aims to go beyond the classification by incorporating novel detection methods that provide detailed information on the occurrence, duration, and frequency of wheeze events accruing in a respiratory record. This detection method can allow for personalized treatment for individuals with respiratory disorders. Furthermore, integrating a detection algorithm into an electronic stethoscope provides real-time alerts of wheeze events present during medical examinations, resulting in more accurate diagnosis and improved care for patients with respiratory disorders.

The paper structure is as follows: Section 2 provides a comprehensive literature review of studies conducted using the ICBHI 2017 open-source data for adventitious sound classification. Section 3 details the methodology followed for building a wheeze detection model. Section 4 presents the results obtained, including the performance matrices and visual representations of them, in addition to a discussion of those results and a comparison between different models and labeling techniques. Finally, Section 5 encapsulates the conclusion of the study.

## 2. Literature Survey

Previous studies have employed different methods that have been applied for respiratory sound classification using the ICBHI 2017 dataset. In several studies, the ICBHI open dataset from 2017 has been a valuable resource [13]. Jakovljevic et al. applied a hidden Markov model to analyze Mel-frequency cepstral coefficients (MFCC) in respiratory sound data [14]. Chambres et al. took a different approach, combining low-level features and MFCC while using a decision tree to identify adventitious lung sounds, achieving an accuracy of 49.63% [15]. Kochetov et al. introduced a classification system for lung sounds, employing a noise-masking recurrent neural network (NMRNN) and utilizing STFT for feature extraction [16]. Ma et al. brought in spectral analysis techniques such as STFT and Wavelet Transform

(WT), implementing a bilinear ResNet (bi-ResNet) neural network to classify respiratory sounds with an accuracy of 52.79% [17]. Ngo et al. ventured into the use of Gamatongue spectrograms, employing an ensemble of Clustering deep neural networks (C-DNN) and Autoencoder networks to classify respiratory cycle anomalies [18]. Acharya et al. proposed a hybrid CNN-RNN model for classifying features derived from Mel spectrograms [19]. Serbes et al. used STFT and WT for feature extraction, employing a support vector machine (SVM) as their classifier, resulting in an accuracy of 49.86% [20]. Demir et al. used the STFT for feature extraction and presented two different deep-learning approaches. The first method included a 16-layer deep convolutional neural network (VGG-16 CNN) with an SVM classifier, achieving an accuracy of 65.5%. In the second approach, transfer learning was employed with the CNN model and a softmax function, yielding an accuracy of 63.09% [21]. In 2020, Demir et al. introduced a CNN model trained using spectrogram images paired with a Linear Discriminant Analysis (LDA) classifier and the Random Subspace Ensembles (RSE) method, resulting in an accuracy of 71.15% [22]. Bilal M. devised a system focused on extracting spectrograms from lung sound signals and feeding them into a custom 12-layer CNN, achieving an accuracy of 64.5% [23].

Asatani et al. employed STFT for extracting the features from the data and integrated Convolutional Recurrent Neural Networks (CRNNs) with bidirectional Long Short-Term Memory (bi-LSTM) blocks to enhance classification accuracy [24]. Similarly, Yang et al. employed STFT for feature extraction and developed a framework using ResNet-18, incorporating Squeeze-and-Excitation (SE) and Spatial Attention Blocks (SA) [25]. In a related vein, Liu et al. classified adventitious respiratory sounds, using Log Mel-filterbank (LMFB) for feature extraction and employing CNN for classification, achieving an accuracy of 81.62% [26]. Likewise, Perna and Tagarelli used MFCCs for feature extraction and recurrent neural networks (RNN) for adventitious respiratory sound classification [27]. Zulfiqar et al. used spectrograms with Artificial Noise Addition (ANA) for feature extraction, employing various CNN architectures (AlexNet, ResNet50, VGG16, and Baseline) for the classification of diverse adventitious respiratory sounds [28]. Similarly, Nguyen and Pernkopf incorporated STFT and Log-Mel for feature extraction, using a pre-trained ResNet model to classify adventitious lung sounds [29]. Ntalampiras and Potamitis introduced a unique feature set based on wavelet analysis, implementing a directed acyclic graph (DAG) network architecture comprising hidden Markov models (HMM) to model their distribution [31]. So far, most research concerning ICBHI data has focused on classifying respiratory sounds. This involves training a machine learning model to identify the presence or absence of certain classes but does not provide details about where these classes are located. Detection combines classification and localization, offering information about the type of object present and its specific location. In our previous study [34], we developed a method for wheeze detection using STFT for feature extraction, and CNN-GRU deep learning model. The built wheeze detection model achieved an F1 score of 0.73. This study is built upon the previous work, however, in this study, an alternative labeling method and other RNN architectures were employed to improve wheeze detection.

Wheeze detection offers unique advantages compared to sound classification alone. In addition to identifying the wheezing event occurrence, this method predicts the duration and frequency of wheezing events throughout the diagnostic process. This leads to tailored treatments and individualized care, enabling healthcare providers to customize interventions and medications for individuals.

## 3. Materials and methods

In this section, the methods employed for constructing a detection model are discussed. Figure 1. provides a visual representation of the methodology workflow.

### 3.1 Data Description

For this study, the public dataset of the 2017 ICBHI competition was used, containing a total of 920 lung sounds recorded by two research teams in Portugal and Greece. The recordings ranged in length from 10 to 90 seconds with sampling frequencies of 4 kHz, 10 kHz, and 44.1 kHz. The cycles can be classified as crackle, wheeze, both crackle, and wheeze, or no ambient noise [13]. Wheezes and crackles are annotated in the dataset with detailed start and end times. The first column in the detailed annotation indicates the start time, the second indicates the end time, and the third indicates the name. Text files with detailed events were used for the study.

### 3.2 Signal Processing

The ICBHI database contains recordings with sampling frequencies of 4 kHz, 10 kHz, and 44.1 kHz. To maintain consistency and compatibility across audio files, they were resampled to a 4 kHz sampling rate. Afterward, to reduce unwanted noise arising from different sources like coughing, intestinal/cardiac sounds, and stethoscope movement, a 12th-order Butterworth band-pass filter with cutoff frequencies set at 120 and 1800 Hz was applied to the recordings.

### 3.3 Preparing the Data for Training

According to the annotations, among a total of 920 audio files, only 341 of them contain one or more instances of wheezing events, totaling 1879 wheeze events in those files. The duration of wheeze segments varies between 0.03 seconds and 5.80 seconds.
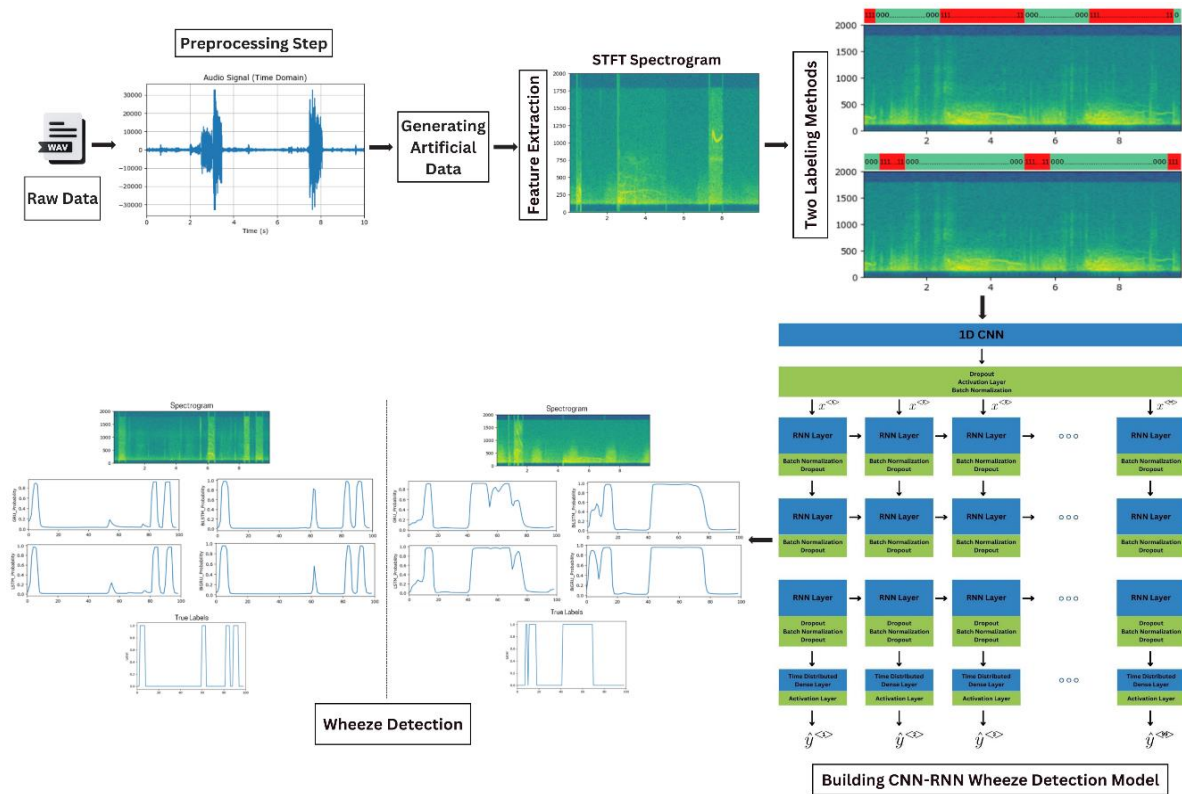
Figure 1. The flowchart of detection framework. The figure contains a visualization of the CRNN model used in the detection process

Figure 2. shows the histogram of the durations, where it can be seen that the majority of wheezes are between the interval of 0.1 seconds and 0.4 seconds. The range of the highest bar, which represents the most frequent durations, in the histogram is between 0.17 and 0.31 seconds. Those numbers match with the literature, which states that wheeze sounds last more than 100 ms on average.

The used dataset encompasses approximately 5.5 hours of recordings as mentioned before. However, when considering the wheezes specifically, the total duration amounts to only approximately 18 minutes of the data. Therefore, the duration of wheezing is only 5.61% of the overall duration of the provided data. This scarcity of wheezing occurrences in the dataset presents a considerable obstacle to developing a reliable wheeze detection model. In addition, during data review, inaccurate event labels in the annotation file were noticed, also, some audio files remained heavily distorted by excessive noise, even after applying the bandpass filter, leading to the emergence of oscillatory patterns that exhibit characteristics similar to wheezing, i.e., exhibiting characteristics like speech or motion artifacts. Consequently, the model may recognize them as instances of wheezing, or it can categorize wheezing as a normal sound. To overcome these limitations, an artificial data generation technique was employed. By creating synthetic data using the annotation files, it is possible to expand the existing dataset and obtain a larger and more diverse set of wheeze events with reduced noise. This strategy aims to enhance the precision of the models by furnishing a more comprehensive training dataset.

### 3.4 Generating Artificial Data

The data was initially split into training and testing sets with a 70% to 30% ratio. Wheezing-containing audio files were identified, and wheezing segments were extracted from them for both training and testing datasets. Figure 3. illustrates the waveform of one of the wheeze segments, where the oscillatory behavior of the wheeze can be clearly seen.

Audio files without wheeze events were also identified, and their first 10 seconds were downloaded. Before using them as background for synthetic data, each file underwent a review, with those containing talking or excessive ambient sounds excluded.

Using wheeze segments and background audio, synthetic data was created. The code randomly selected 10-second backgrounds and added 2 to 4 wheezes to each. 130 different
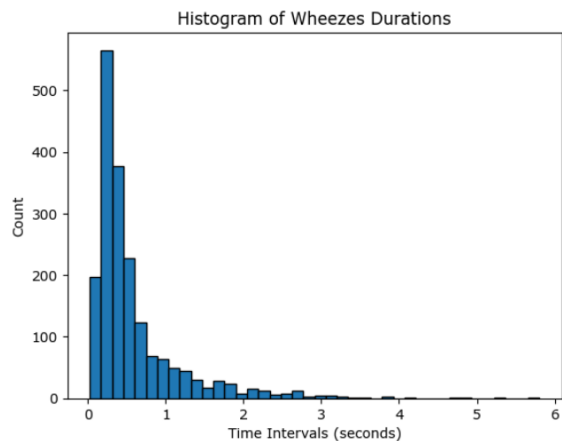


Figure 2. The Histogram of Wheeze Durations

backgrounds with lower noise levels and minimal talking sounds were selected. Among these, 100 background signals were employed for training, while 30 background signals were used for testing.

The code ensures no overlapping segments in selected audio clips to avoid simultaneous wheeze events. It creates 2, 3, or 4 wheeze segments in 10-second sound files. Having at least two wheeze events per audio is crucial to balance the skewed labels (mostly 0s), reducing the risk of overfitting. This approach generates 1500 training and 200 testing audio files.

### 3.5 Short-Time Fourier Transform Spectrogram

The Short-Time Fourier Transform (STFT) is a commonly used method for the time-frequency analysis attributes of a signal. It involves dividing the input signal into overlapping windows and subsequently applying the Fast Fourier Transform (FFT) to each of these windows. The STFT shows how the frequency content of the signal changes over time, making it a powerful tool for analyzing time-varying spectral features. The mathematical representation of the STFT, X[$m,w$], is as follows [32]:

$$X[m,w] = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-iwn} \qquad (1)$$

Where in Equation (1), $x[n]$ represents the input signal, $w[n]$ refers to the window function and N is the length of the FFT. Squaring the magnitude of the STFT, $|X[m,w]|^2$, gives the spectrogram which is the visual representation of the signal in the time-frequency domain.

When visualizing the audio signal's spectrogram, it becomes evident that specific patterns associated with the characteristic frequency range become clearly visible. For the best visualization of the patterns, and for obtaining accurate features to be fed to the model, the parameters of the STFT should be selected carefully by giving attention to both the time resolution and the frequency resolution of the spectrogram. A narrow window provides good time resolution but sacrifices frequency resolution, while a wide window offers poor time resolution but good frequency resolution. Since wheezes last a long time and have a narrow frequency pattern, there is a need for better frequency resolution to visualize them in the time-frequency domain.

Figure 4 displays a plot of an audio signal alongside its corresponding spectrogram. In this study, a Blackman-Harris (BH) window was used for wheeze visualization with a length of 512 and a 192 overlap (75% of the window size). This allows us to clearly see wheezes as narrowband spectral patterns, often referred to as "snakes" in the spectrogram.

### 3.6 Labeling the Data

The labels were produced using two distinct approaches. In the first approach, the labels for the audio signals were synchronized with the time frames of the spectrograms, assigning one label for every three-time steps. The portions
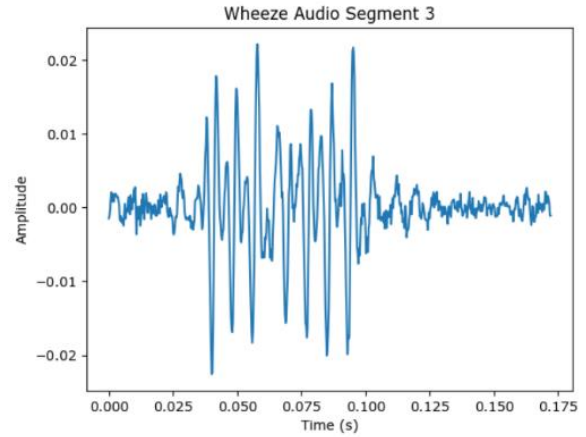


Figure 3. Wheeze segment plot in the time domain
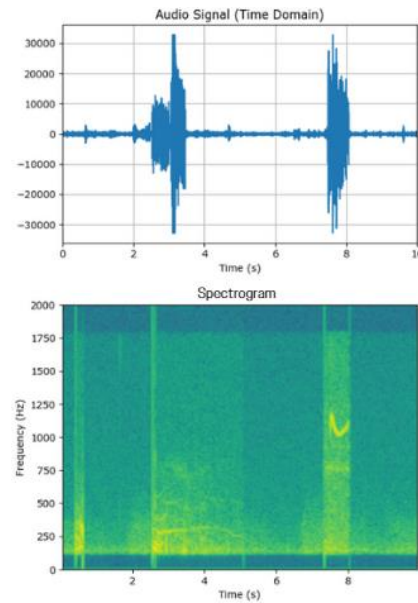


Figure 4. Audio signal plot in the time domain and its corresponding STFT spectrogram

of the audio that lack wheezes are assigned zero. While labeling the portions of wheezing, a specific approach was employed.

In this approach, a target label of 1 is assigned to the 10 consecutive time steps following the end of a wheeze-containing clip. The goal of labeling in this way is to train the model to recognize when wheezing sounds show up, with a focus on the time immediately after the clip ends. Each generated spectrogram has a label in the shape of (1, 99) which is the output of the model. The labeling of segments is visually represented in Figure 5A, where a binary classification is applied: "the time steps immediately after the wheezing event shows up" (labeled as 1) or " time steps without wheezes, or wheezing events" (labeled as 0).

In the second approach, the labels are assigned directly to the time steps where the wheeze occurs. In this way, the model is supposed to be able to predict the exact time of the

wheeze appearance. Similar to the first approach, the shape of the labels is (1, 99). The labeling of segments is visually represented in Figure 5B, where a binary classification is applied: "with wheeze" (labeled as 1) or "without wheeze" (labeled as 0).

### 3.6 RNN-Based Model Architecture

In this study, four RNN-based models, namely CNN-LSTM, CNN-GRU, CNN-BiLSTM, and CNN-BiGRU, were investigated. The built model architecture includes a single convolutional layer, followed by three RNN layers, and finalized with a dense layer.

The model starts with a convolutional layer, taking a 309-time step input. The convolutional layer is critical to extract low-level features and decrease the output dimensionality. This layer has 196 units, uses a 15-sized kernel, and has a stride of 3. This adjustment aligns the model's output dimensionality with the label dimensions (99). The CNN layer accelerates learning by reducing input dimensionality. Consequently, batch normalization is applied to normalize the feature, and then the ReLU activation function is used to introduce non-linearity. A dropout layer is added to prevent overfitting.

Following the CNN layer, RNN layers are employed. The architecture of the four models is identical in terms of the number of layers, dropouts, batch normalizations, and other parameters. The only difference lies in the type of RNN blocks used. The first layer utilizes 512 units and return sequences, allowing the model to capture temporal relationships and patterns within the dataset. Dropout and batch normalization are applied to enhance the stability of the model. The next RNN layer also utilizes 512 units. This layer further captures complex temporal relationships in the data. Dropout and batch normalization are applied similarly to the previous layer. The last RNN layer employs 256 units and serves to prepare the data for the subsequent output layer. Dropout and batch normalization are applied again to ensure robustness and prevent overfitting. The final layer consists of a time-distributed dense layer with sigmoid activation. The model's architecture is presented in Table 1.

The model was trained for 30 epochs. For the evaluation of the models, several metrics were generated to assess their performance. These metrics include accuracy, precision, recall, and F1 score.

## 4. Results and Discussion

### 4.1 Models' Accuracies

Table 2 presents the evaluation metrics for the model, comparing the two different labeling techniques. The CRNN models with positive labels starting after the events achieved remarkably similar accuracies ranging between 0.82 and 0.84. Notably, the GRU-based model achieved the lowest accuracy, while the bidirectional models exhibited the highest accuracy among the four models. These accuracies

indicate that the model can correctly identify wheeze patterns in 82-84% of cases. Similarly, the CRNN models with the positive labels aligned directly on the wheeze events demonstrated accuracies within the same range as the models using the alternative labeling technique. However, in this case, the LSTM-based model achieved the lowest accuracy of 0.82.

While accuracy is a crucial metric for classification tasks, in scenarios where the distribution of labels is heavily skewed towards the "0" class, accuracy may not provide the most informative assessment. In such cases, defining more meaningful metrics such as F1 score, Precision, and Recall, offers a more insightful evaluation of the model's performance. By considering those metrics, a better understanding of the model's ability to correctly predict positive class while minimizing false positives and false negatives can be achieved.

### 4.2 Models' Precision and Recall Values

Precision evaluates how effectively the model can correctly identify positive predictions. Equation (2) represents the mathematical calculation of precision. A higher precision score signifies the model's capacity to correctly identify wheezing patterns while reducing the misclassification of non-wheezing events as wheezing.

$$Precision = \frac{\sum TP}{\sum (TP + FP)} \qquad (2)$$

Recall assesses the model's capacity to accurately identify all positive instances. Equation (3) represents the mathematical calculation of the recall. The recall score evaluates the model's ability to detect the real instances of wheezing in the dataset, ensuring that it doesn't overlook any existing wheezing events.

$$Recall = \frac{\sum TP}{\sum (TP + FN)} \qquad (3)$$

Table 1. Model Architecture

| Layer Type | Kernel Attribute | Activation |
|---|---|---|
| Conv1D | 15 (196 Filters) | ReLu |
| BatchNormalization | | - |
| Dropout | 0.8 | - |
| RNN | 512 units | - |
| Dropout | 0.8 | - |
| BatchNormalization | | - |
| RNN | 512 units | - |
| Dropout | 0.8 | - |
| BatchNormalization | | - |
| RNN | 256 units | - |
| Dropout | 0.8 | - |
| BatchNormalization | | - |
| Dropout | 0.8 | - |
| TimeDistributed | Dense 1 | Sigmoid |
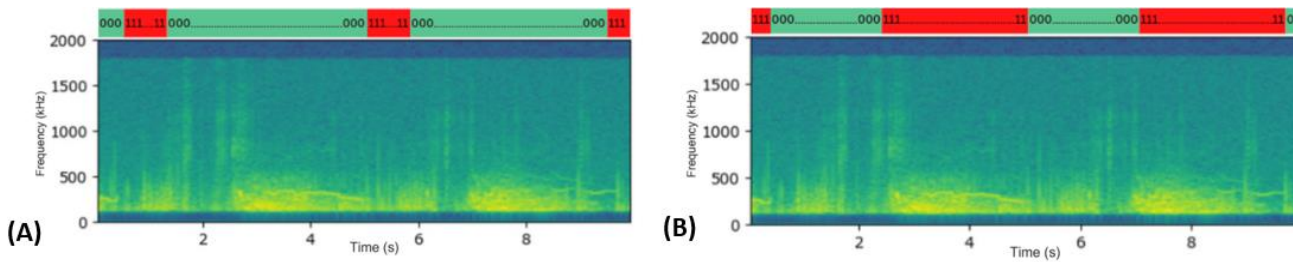
Figure 5. Visual Representation of the Labeling. (A) Positive Labeling After Wheeze Events and (B) Positive Labeling Directly on Wheeze Events

Table 2. Accuracy, Precision, Recall, and F1 Score Values

| Metric | | Positive Labels After Wheeze Events | | | | Positive Labels on Wheeze Events | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | CNN-LSTM | CNN-GRU | CNN-BLSTM | CNN-BGRU | CNN-LSTM | CNN-GRU | CNN-BLSTM | CNN-BGRU |
| Accuracy | | 0.83 | 0.82 | 0.84 | 0.84 | 0.82 | 0.83 | 0.84 | 0.84 |
| Precision | 0 | 0.86 | 0.87 | 0.87 | 0.85 | 0.85 | 0.85 | 0.86 | 0.86 |
| | 1 | 0.75 | 0.60 | 0.67 | 0.78 | 0.73 | 0.75 | 0.75 | 0.79 |
| Recall | 0 | 0.95 | 0.90 | 0.92 | 0.96 | 0.92 | 0.93 | 0.92 | 0.94 |
| | 1 | 0.47 | 0.54 | 0.53 | 0.43 | 0.57 | 0.56 | 0.61 | 0.58 |
| F1 score | 0 | 0.89 | 0.88 | 0.90 | 0.91 | 0.88 | 0.89 | 0.89 | 0.90 |
| | 1 | 0.58 | 0.57 | 0.59 | 0.56 | 0.64 | 0.64 | 0.68 | 0.67 |
| Macro average | Precision | 0.80 | 0.74 | 0.77 | 0.82 | 0.79 | 0.80 | 0.81 | 0.82 |
| | Recall | 0.71 | 0.72 | 0.73 | 0.70 | 0.74 | 0.74 | 0.77 | 0.76 |
| | F1 Score | 0.74 | 0.73 | 0.74 | 0.73 | 0.76 | 0.76 | 0.78 | 0.78 |

Examining the precision values in Table 2, it becomes evident that almost all models supplied with positive labels after wheeze events achieve relatively high values for negative class (0) which refers to wheeze absence, ranging from 0.86 to 0.87. This indicates the ability of the models to correctly classify non-wheezing events. However, the precision for the positive classes (1) which refers to the wheeze events, varies between the models. The CNN-BiGRU model achieves the highest precision of 0.78, while the CNN-GRU model has the lowest precision of 0.60. The macro average represents the average across both classes, treating them equally. The macro precision average ranges between 0.74 and 0.82, with the highest value for the CNN-

BiGRU model. These results indicate the models can precisely classify approximately 74%-82% of the positive instances. Regarding recall of the same labeling technique, the models show varying performances. For the negative class (0), the recall values range between 0.90 and 0.96, suggesting the models' effectiveness in capturing true negative (TN) instances.

However, for the positive class (1), the recall values range from 0.43 to 0.54, indicating that the model's ability to capture positive wheeze events ranges between 43% and 54%. The macro average of the models ranges between 0.70 to 0.73, with the highest value for the CNN-BiLSTM model. This indicates that the models exhibit a balanced

performance in terms of correctly identifying wheezing events and non-wheezing events.

When examining the precision and recall values derived from the second labeling technique, which directly labels wheeze events as positive, positive differences in performance become apparent when compared to the first labeling technique within the same table. In terms of precision, all the models achieve relatively high values for both the negative class (0) and the positive class (1) when compared with the alternative labeling technique. The precision value values range between 0.85 and 0.86 for the negative class and from 0.73 to 0.79 for the positive class.

For recall, the models again exhibit varying performance with relatively high values for the negative class. However, the recall values for the positive class, in the second technique that directly labels wheeze events as positive, range from 0.56 to 0.61, suggesting that the models have varying degrees of success in identifying wheezing events. Overall, the recall values are relatively higher than the alternative model. The macro average of the models ranges between 0.74 to 0.77, with the highest value for the CNN-BiLSTM model.

### 4.3 Models' F1 Scores

The F1 score is calculated as the harmonic mean of precision and recall. It is a metric that combines precision and recalls into a single value providing a balanced measure of models' performance. Considering the unbalanced nature of the labels, where the positive class (wheeze events) is a minority class, it is crucial to focus on the F1 score of the positive class (1). Looking at Table 2, it can be observed that among the models, CNN-BiLSTM and CNN-BiGRU consistently demonstrate higher F1 scores for wheeze events across both labeling techniques. For positive labeling after wheeze events, the BiLSTM model achieves the highest F1 score of 0.59 for the positive labels, while for labeling directly on wheeze events both bidirectional models give the same highest F1 score of 0.68. Comparing the F1 scores for the positive class, the second labeling technique outperforms the models with positive labeling after the wheeze events. This suggests that labeling directly on wheeze events provides a more balanced correct identification of wheeze events (precision) and captures all actual wheeze events (recall).

In order to provide a more comprehensive overview of the model's performance, two bar charts were illustrated to visualize the accuracy and the macro averages of the precision, recall, and F1 score. Figure 6 shows the scores of the models with positive labels after wheeze events, while Figure 7 depicts the scores of the models with positive labels on wheeze events. These charts provide a concise and visual representation of the data presented in Table 2. Upon closer inspection of the figures, it becomes evident that the bidirectional models outperform other architectures in both

labeling techniques.

### 4.4 Visualization of the Obtained Results

In this study, results obtained from the models are visualized to offer a visual assessment of the model's performance in detecting wheezing events. Both the spectrogram and the prediction values were used to assess the model's ability in this regard. In this section, two examples are presented. The first example is from the artificial test data, and the second is from one of the original audios of ICBHI data. Figure 8 illustrates the results of one created audio example with positive labeling after the wheeze event. Peaks or changes in the prediction values of the probability graphic show when the model predicts the occurrence of a wheeze event. It can be seen that all the models except the BiGRU model could predict all the wheezes in the audio. GRU model's prediction values were less than 0.8, whereas the LSTM-based models gave prediction values of more than 0.9. Figure 9 illustrates the results of the same example but with the positive labels aligned directly on the wheezes. Looking at the true labels it can be seen that the bidirectional models could identify all the present wheeze events in the audio.
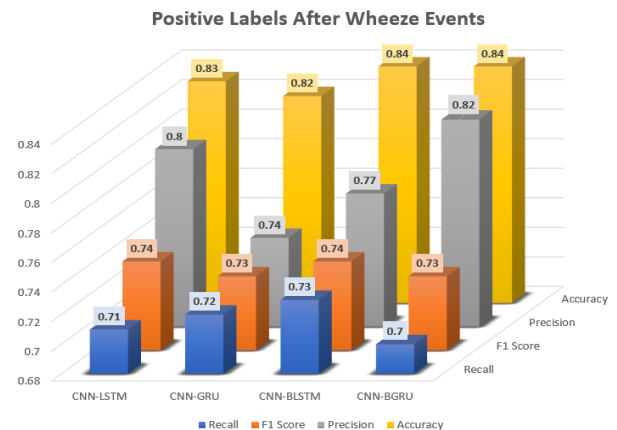


Figure 6. Accuracy, precision, F1 score, and recall scores for models with positive labels after wheeze events
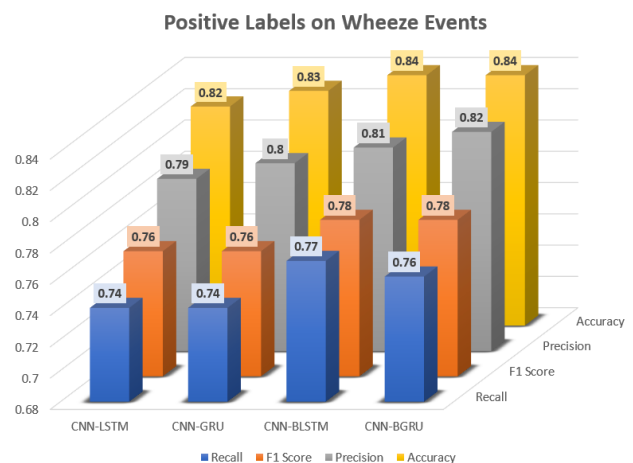


Figure 7. Accuracy, precision, F1 score, and recall scores for models with positive labels on wheeze events
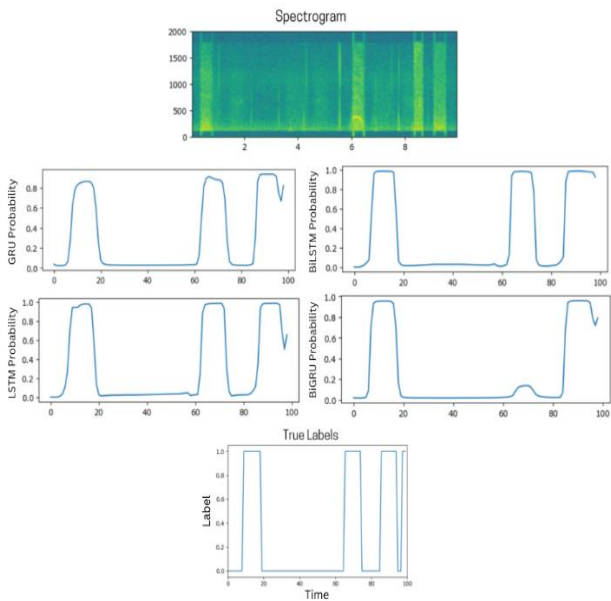
Figure 8. Visualization of the spectrogram, the predictions, and the true labels with positive labels after the wheeze events.
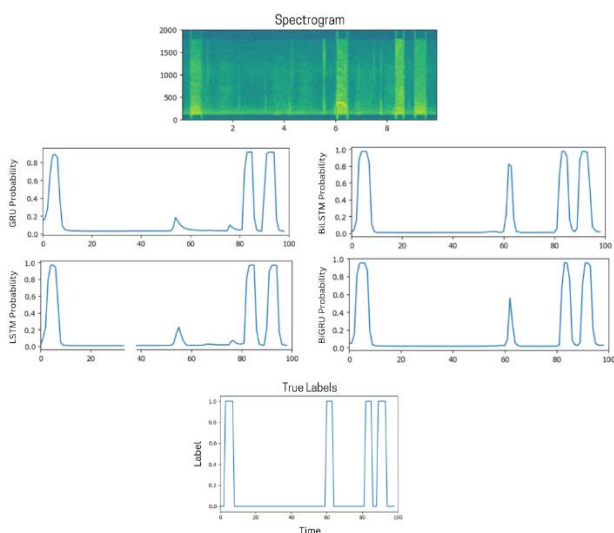


Figure 9. Visualization of the spectrogram, the predictions, and the true labels with positive labels aligned on the wheeze events



Annotation
['0.036'　'0.393'　'wheeze']
['2.564'　'5.007'　'wheeze']
['6.921'　'9.493'　'wheeze']
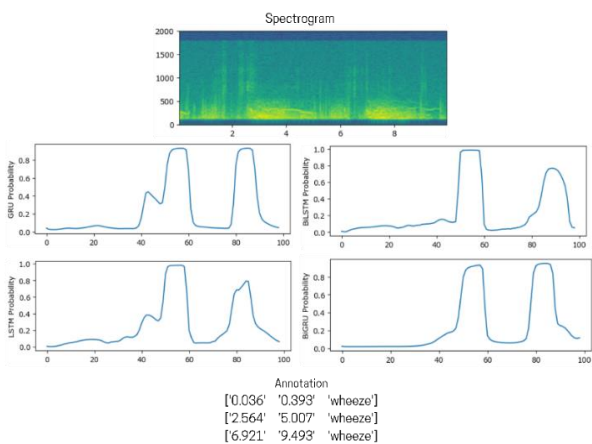
Figure 10. Visualization of the spectrogram, the predictions, and the true labels with positive labels after the wheeze events



Annotation
['0.036'　'0.393'　'wheeze']
['2.564'　'5.007'　'wheeze']
['6.921'　'9.493'　'wheeze']
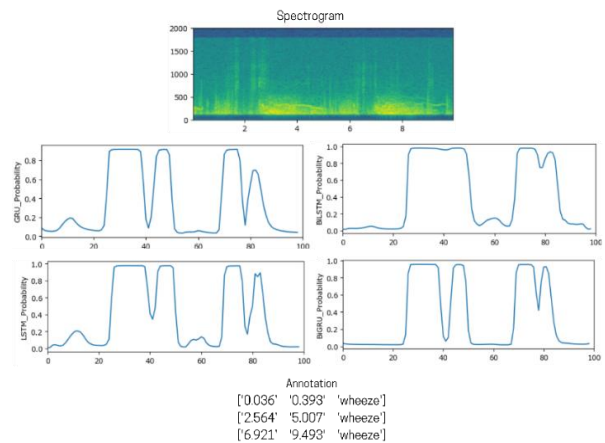
Figure 11. Visualization of the spectrogram, the predictions, and the true labels with positive labels aligned on the wheeze events.

The BiGRU model could detect the wheeze after the second 6 with a prediction value of 0.6, whereas the BiLSTM model was able to detect the same wheeze event with a prediction value of 0.8. On the other hand, the unidirectional models were not able to detect it. Figure 10 illustrates the results of one of the original audio examples with positive labeling after the wheeze events. From the annotation, it can be seen that all the models were not able to detect the short wheeze event at the duration between 0.036 and 0.393, whereas all of them were able to predict the other two events. Figure 11 illustrates the results of the same audio, but with labels directly assigned to the wheeze events. Looking at the annotations, it can be noticed that all models were not able to predict the same wheeze event at the start of the audio, however, for the other wheeze events the bidirectional models gave a more accurate prediction, with the best prediction given from the BiLSTM model. A set of other synthetic data and original data were also visualized, and as a visual result, it was noticed that the models with positive labels assigned directly to the wheeze events, especially the BiLSTM model, were able to detect the wheeze events more precisely.

### 4.5 Comparison Between Models

The conclusive comparison between LSTM-based and GRU-based models is difficult, however, in terms of computation times, the GRU-based models have been proven to be faster than the LSTM models [33]. As mentioned before, with the labels being unbalanced, and the positive class (wheeze events) being a minority, it is crucial to focus on the F1 score of the positive class (1). Therefore, focusing on the positive class of the F1 score it can be noticed that the LSTM-based models outperform other models in the first labeling technique. This difference in performance could be explained by the LSTM architecture, which includes a memory cell and gating mechanisms that allow it

to capture and remember long-term dependencies and retain information over a longer sequence. In the case of wheeze detection, where multiple time steps may be required to detect wheezing, these long-term dependencies are particularly useful. TN and TP values shown previously may be slightly lower in GRU models due to their relatively weaker ability to retain and utilize longer-term information.

On the other hand, in positive labeling directly on wheeze events, looking at the F1 scores of positive labels, it can be noticed that the bidirectional models tend to outperform the unidirectional models. Bidirectional models have the advantage of considering both past and future contexts during training. The reason for this can be that bidirectional models are capable of capturing bidirectional dependencies between wheezing events. Using bidirectional models, input sequences can be processed in both forward and backward directions, providing a more comprehensive input understanding.

Looking at the tables above, it is noticed that in both labeling techniques, the CNN-BiLSTM model achieves the highest F1 score for positive labels and the macro average. This can be attributed to the advantages offered by its architecture which combines the strength of both LSTM and bidirectional models. Therefore, the CNN-BiLSTM model with its combination of long-term dependencies capture and remember from LSTM and bidirectional context modeling, offers a favorable balance between capturing temporal patterns of wheeze events and incorporating bidirectional dependencies, making it well-suited for the wheeze detection task.

### 4.6 Comparison Between Labeling Techniques

In this study, two different labeling techniques were used. The results of each technique were obtained separately and presented in previous parts. The first labeling approach involves assigning positive labels following wheeze events. By applying this labeling method, the models consider the entire context of the wheeze events' pattern, allowing for a thorough analysis of those patterns. This labeling approach is generally applied in speech recognition tasks, where the labeling is performed after the speech utterance. However, in contrast to speech recognition or trigger word detection scenarios, the duration of spoken words is typically constant, whereas the lengths of wheeze events in the dataset vary significantly. This variability in wheeze duration presents a challenge in labeling wheezes after the event, as a fixed number of positive labels may not adequately capture the varying lengths of wheezes present in the data set. This challenge suggests relatively lower F1 scores for positive labels compared with the alternative method. Implementing this labeling technique along with using the CNN-LSTM model for wheeze detection offers the advantage of real-time alarm generation. By implementing the algorithm into a stethoscope, it may be possible to receive an instant alert

when a wheeze occurs. This instantaneous feedback can be valuable in medical settings to tackle the problems mentioned in the introduction section of this study. However, using bidirectional models with this labeling method requires waiting for the recording of respiratory sounds to finish, which limits the real-time capabilities.

Positive labeling directly aligned with the wheeze event offers several advantages. Firstly, more precise identification of wheeze occurrence, allowing for accurate duration estimation, can be achieved. Furthermore, direct labeling in the event can capture fine-grained details of the wheezing pattern, leading to a higher F1 score for positive labels. It enables the model to learn specific wheeze features, resulting in enhanced differentiation between wheeze and non-wheeze segments. To implement the CNN-LSTM system with such a labeling approach to an electronic stethoscope for wheeze detection to provide an instant alarm, it typically needs to wait for the wheeze pattern to reach a recognizable state. While the system may not be able to provide an instant alarm at the exact onset of the wheeze, once a sufficient portion of the wheeze event is detected and recognized, the system can trigger an alarm to indicate the presence of a wheeze.

In summary, even though the first labeling technique allows comprehensive analysis of the wheeze pattern, the variable lengths of wheeze events can be challenging and result in lower F1 scores for positive labels. On the other hand, labeling directly aligned with the event provides precise identification and duration estimation of wheeze occurrence, yielding higher F1 scores.

### 4.7 Comparison with Other Studies

Table 3 shows some of the most recent and relevant studies in the literature for the ICBHI 2017 database. The table contains the studies that utilize STFT or comparable time-frequency representations, such as Mel spectrograms or MFCCs, and employ deep learning-based methods for classification. In the table, the column "parameters" denotes the window length used for time-frequency representation. Furthermore, columns "Sen", "Spe", "Sco", and "Acc" refer to the ICBHI 2017 benchmark sensitivity, specificity, scoring result, and classification accuracy, respectively. Although the studies listed in Table 3 employed methods similar to those utilized in this study, it's crucial to note a distinction. These studies focused on classifying abnormal respiratory sounds, and training models to identify the presence or absence of adventitious lung sounds in the record. However, they lacked the capacity to provide localization information about these classes. As mentioned previously, the majority of works on the ICBHI 2017 database have centered around classifying respiratory sounds without offering details about where these sounds occur. In contrast, our study stands out by aiming to develop a model for wheeze detection, which integrates both the classification and the localization of the wheeze sounds.

Table 3. Studies in Literature for the ICBHI 2017 Dataset.  Column "parameters" denotes the window length for employed time-frequency representation.

| Reference | Time-frequency representation | | Learning model | Results (%) | | | |
|---|---|---|---|---|---|---|---|
| | Type | parameters | | Sen | Spe | Sco | Acc |
| Kochetov et al. [16] | STFT | 500 ms | RNN | 58.4 | 73 | 65.7 | - |
| Acharya et al. [19] | Mel spectrograms | 60 ms | hybrid CNN-RNN | - | 58.01 | - | - |
| Liu et al. [26] | LMFB | - | CNN | - | - | - | 81.62 |
| Asatani et al. [24] | STFT | 40 ms | CRNN, bi-LSTM | 63 | 83 | 73 | - |
| Perna &Tagarelli [27] | MFCCs | 250 ms | RNN | 64 | 84 | 74 | - |
| Saraiva et al. [30] | MFCCs | - | CNN | - | - | - | 74.3 |

## 5. Conclusions

Chronic respiratory disorders have a significant impact on lung function and overall health, with conditions such as chronic obstructive pulmonary disease and asthma contributing to a high number of deaths globally. Early diagnosis and regular monitoring of respiratory illnesses are crucial for effective management. While traditional auscultation using a stethoscope has been a common diagnostic tool, it has limitations in terms of subjectivity and accuracy. Computer-based respiratory sound analysis has emerged as a promising approach to improve the objectivity and efficiency of diagnosis.

The objective of this study is to develop a computer-based system capable of detecting wheeze sounds, to address the limitations of the traditional auscultation devices mentioned before. By implementing computer-based respiratory sound analysis using artificial data derived from ICBHI open data, conventional recurrent (CRNN) models were trained to detect wheeze in the created respiratory recordings.

In the implementation of the wheeze detection model, one of the major challenges faced was the limited open data of respiratory sound recordings. The available open data that was used in this project is relatively small, comprising only 920 audio files, of which only 341 files contain one or more wheeze events. In this study, to tackle this limitation, artificial data was generated from the ICBHI open data. However, having a larger dataset with a higher number of patients, a higher number of wheeze events, and reduced noise would significantly enhance the performance of the wheeze detection models, leading to more efficient and accurate wheeze detection capabilities.

Before training the data, two different methods of labeling were applied to the created recordings, each with its advantages and considerations. Those techniques are; positive labeling after wheeze events, and positive labeling aligned with the wheeze events. The second labeling method showed more precise identification, duration estimation, and enhanced differentiation between wheeze and non-wheeze segments.

The comparison between different models and labeling techniques showed that the LSTM-based models demonstrated better performance in terms of F1 score for positive labels after the wheeze events. On the other hand, the bidirectional models showed superior performance and F1 scores for positive labeling directly aligned with wheeze events. The CNN-BiLSTM model emerged as the most effective model, leveraging the strengths of both LSTM and bidirectional architecture. It combines the ability to capture long-term dependencies and incorporate bidirectional dependencies making it suitable for wheeze detection problems.

In future studies, it is recommended to address the limited dataset issue by acquiring a larger dataset with more patients, wheezing events, and reduced noise. Moreover, using K-fold cross-validation can lead to a more robust analysis, enhancing the reliability of wheeze detection models. This would further improve the efficiency and accuracy of computer-based respiratory sound analysis in order to foster their potential for early diagnostics and monitoring of respiratory diseases.

## Declaration

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

## Author Contributions

L. Hakki and G. Serbes developed the methodology, performed the analysis, and wrote the manuscript together. G. Serbes proofread the manuscript.

## Nomenclature

ICBHI   : International Conference on Biomedical and Health Informatics
CRDs    : Chronic Respiratory Disorders
COPD   : Chronic Obstructive Pulmonary Disease
CNN     : Convolutional Neural Network
RNN     : Recurrent Neural Network
NMRNN: Noise Masking Recurrent Neural Network
ANA     : Artificial Noise Addition
HMM    : Hidden Markov Model
SVM     : Support Vector Machine
DAG     : Directed Acyclic Graph
LDA     : Linear Discriminant Analysis
RSA     : Random Subspace Ensembles
SE       : Squeeze-and-Excitation
SA       : Spatial Attention Block
STFT    : Short-Time Fourier Transform
WT       : Wavelet Transform
LMFB   : Log Mel-Filterbank
MFCC  : Mel-Frequency Cepstrum Coefficient
GRU     : Gated Recurrent Unit
LSTM   : Long Short-Term Memory
Bi-LSTM: Bidirectional Long Short-Term Memory
Bi-GRU : Bidirectional Gated Recurrent Unit
TP       : True Positive
TN      : True Negative
FP       : False Positive
FN      : False Negative
Hz      : Hertz
ms     : milliseconds

## References

1. Cukic, V., Lovre, V., Dragisic, D., & Ustamujic, A. *Asthma and chronic obstructive pulmonary disease (COPD) – differences and similarities.* Materia Socio-Medica, 2012. **24**(2): p. 100.

2. World Health Organization. (n.d.). Chronic obstructive pulmonary disease (COPD). World Health Organization. Retrieved [cited October 25, 2022]; Available from: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd).

3. Liang, R., Feng, X., Shi, D., Yang, M., Yu, L., Liu, W., Zhou, M., Wang, X., Qiu, W., Fan, L., Wang, B., & Chen, W. *The global burden of disease attributable to high fasting plasma glucose in 204 countries and territories, 1990-2019: An updated analysis for the Global Burden of Disease Study 2019.* Diabetes/metabolism research and reviews, 2022. **38**(8): e3572.

4. Göğüş, F. Z., Karlık, B., & Harman, G. *Classification of asthmatic breath sounds by using wavelet transforms and neural networks.* International Journal of Signal Processing Systems, 2014. **3**(2): p. 106-111.

5. Güler, İ., Polat, H., & Ergün, U. *Combining neural network and genetic algorithm for prediction of lung sounds.* Journal of Medical Systems, 2005. **29:** p. 217-231.

6. Yeginer, M., & Kahya, Y. P. *Feature extraction for pulmonary crackle representation via wavelet networks.* Computers in Biology and Medicine, 2009. **39**(8): p. 713–721.

7. Reichert, S., Gass, R., Brandt, C., & Andrès, E. *Analysis of respiratory sounds: State of the art.* Clinical Medicine: Circulatory, Respiratory and Pulmonary Medicine, 2008. p. 45-58.

8. Pasterkamp, H., & Zielinski, D. *The History and Physical Examination.* Kendig's Disorders of the Respiratory Tract in Children, 2019 (9th Edition). p. 2–25.

9. Sarkar, M., Madabhavi, I., Niranjan, N., & Dogra, M. *Auscultation of the respiratory system.* Annals of Thoracic Medicine, 2015. **10**(3): p. 158-168.

10. Zaitseva, E. G., Chernetsky, M. V., & Shevel, N. A. *About Possibility of Remote Diagnostics of the Respiratory System by Auscultation.* Devices and Methods of Measurements, 2020. **11**(2): p. 148-154.

11. Kim, Y., Hyon, Y., Jung, S. S., Lee, S., Yoo, G., Chung, C., & Ha, T. *Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning.* Scientific Reports, 2021. **11**(1): p. 1-11.

12. Hsu, F.-S., Huang, S.-R., Huang, C.-W., Huang, C.-J., Cheng, Y.-R., Chen, C.-C., Hsiao, J., Chen, C.-W., Chen, L.-C., Lai, Y.-C., Hsu, B.-F., Lin, N.-J., Tsai, W.-L., Wu, Y.-L., Tseng, T.-L., Tseng, C.-T., Chen, Y.-T., & Lai, F. *Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—hf_lung_v1.* PLOS ONE, 2021. **16**(7): p. 1-26.

13. Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P., … de Carvalho, P. *An open access database for the evaluation of respiratory sound classification algorithms.* Physiological Measurement. 2019. **40**: 035001

14. Jakovljević, N., & Lončar-Turukalo, T. *Hidden Markov Model Based Respiratory Sound Classification.* IFMBE Proceedings, 2017. **66**: p. 39–43.

15. Chambres, G., Hanna, P., & Desainte-Catherine, M. *Automatic Detection of Patient with Respiratory Diseases Using Lung Sound Analysis.* 2018 International Conference on Content-Based Multimedia Indexing (CBMI). 2018. p. 1-6.

16. Kochetov, K., Putin, E., Balashov, M., Filchenkov, A., & Shalyto, A. 2018. *Noise Masking Recurrent Neural Network for Respiratory Sound Classification.* In Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018. Springer International Publishing. p. 208–217.

17. Ma, Y., Xu, X., Yu, Q., Zhang, Y., Li, Y., Zhao, J., & Wang, G. *LungBRN: A Smart Digital Stethoscope for Detecting Respiratory Disease Using bi-ResNet Deep Learning Algorithm.* 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2019. IEEE. p. 1-4.

18. Ngo, Pham, L., Nguyen, A., Phan, B., Tran, K., & Nguyen, T. (2021). *Deep Learning Framework Applied For Predicting Anomaly of Respiratory Sounds.* 2021 International Symposium on Electrical and Electronics Engineering (ISEE). IEEE. p. 42-47.

19. Acharya, J., & Basu, A. *Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning.* IEEE Transactions on Biomedical Circuits and Systems, 2020. **14**(3): p. 535-544.

20. Serbes, G., Ulukaya, S., & Kahya, Y. P. *An Automated Lung Sound Preprocessing and Classification System Based OnSpectral Analysis Methods.* In Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017. Springer Singapore. p. 45-49.

21. Demir, F., Sengur, A., & Bajaj, V. *Convolutional neural networks based efficient approach for classification of lung diseases.* Health Information Science and Systems, 2019. **8**(1): 4.

22. Demir, F., Ismael, A. M., & Sengur, A. *Classification of lung sounds with CNN model using parallel pooling structure.* IEEE Access, 2020. **8**: p. 105376-105383.

23. ER, M. B. *Akciğer Seslerinin Derin öğrenme ile sınıflandırılması.* Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 2020. **8**(4): p. 830–844. (In Turkish).

24. Asatani, N., Kamiya, T., Mabu, S., & Kido, S. *Classification of respiratory sounds using improved convolutional recurrent neural network.* Computers & Electrical Engineering, 2021. **94**: 107367.

25. Fan, C.-Y., Liu, C.-P., Wang, K.-C., Jhan, J.-H., Wang, Y.-C. F., & Chen, J.-C. *Face Feature Recovery via Temporal Fusion for Person Search.* ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. 2020, IEEE. p. 1893-1897.

26. Liu, R., Cai, S., Zhang, K., & Hu, N. *Detection of Adventitious Respiratory Sounds based on Convolutional Neural Network.* 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). 2019, IEEE. p. 298-303.

27. Perna, D., & Tagarelli, A. *Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks.* 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). 2019, IEEE. p. 50-55.

28. Zulfiqar, R., Majeed, F., Irfan, R., Rauf, H. T., Benkhelifa, E., &amp; Belkacem, A. N. *Abnormal respiratory sounds classification using deep CNN through Artificial Noise addition.* Frontiers in Medicine, 2021. **8**: 714811.

29. Nguyen, T., &amp; Pernkopf, F. *Lung sound classification using co-tuning and stochastic normalization.* IEEE Transactions on Biomedical Engineering, 2022. **69**(9): p. 2872–2882.

30. Saraiva, A., Santos, D., Francisco, A., Sousa, J., Ferreira, N., Soares, S., &amp; Valente, A. *Classification of respiratory sounds with convolutional neural network.* Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies. 2020, Science and Technology Publications. p. 138-144.

31. Ntalampiras, S., & Potamitis, I. *Automatic acoustic identification of respiratory diseases.* Evolving Systems, 2020. **12**(1): p. 69-77.

32. Krishnan, S. *Advanced Analysis of Biomedical Signals.* Biomedical Signal Analysis for Connected Healthcare, 2021: p. 157–222.

33. Li, L., Wu, Z., Xu, M., Meng, H. M., & Cai, L. *Combining CNN and BLSTM to Extract Textual and Acoustic Features for Recognizing Stances in Mandarin Ideological Debate Competition.* In Interspeech, 2016. p. 1392-1396.

34. Hakki, L., & Serbes, G. *Wheeze Events Detection Using Convolutional Recurrent Neural Network.* In 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), Sivas, Turkiye, 2023. IEEE. p. 1-6.

*Research Article*

# A Modified MFCC-based deep learning method for emotion classification from speech

*Fatih Şengül [a]* [ID] *and Sıtkı Akkaya [b,]* [ID]

[a] *Department of Defense Technologies, Institute Of Graduate Studies, Sivas University of Science and Technology, Sivas 58100, Turkey*
[b] *Department of Electrical and Electronics Engineering, Faculty of Engineering and Natural Sciences, Sivas University of Science and Technology, Sivas 58100, Turkey*

| ARTICLE INFO | ABSTRACT |
|---|---|
| | Speech, which is one of the most effective methods of communication, varies according to the emotions experienced by people and includes not only vocabulary but also information about emotions. With developing technologies, human-machine interaction is also improving. Emotional information to be extracted from voice signals is valuable for this interaction. For these reasons, studies on emotion recognition systems are increasing. In this study, sentiment analysis is performed using the Toronto Emotional Speech Set (TESS) created by the University of Toronto. The voice data in the dataset is first preprocessed and then a new CNN-based deep learning method is compared. The voice files in the TESS dataset first yielded feature maps using the Mel Frequency Cepstral Coefficient (MFCC) method, and then classification was performed with this method based on the proposed neural network model. Separate models have been created with CNN and LSTM (Long Short-Term Memory) models for the classification process. The experiments show that the MFCC-applied CNN model achieves a better result with an accuracy of 99.5% than the existing methods for the classification of voice signals. The accuracy value of the CNN model shows that the proposed CNN model can be used for emotion classification from human voice data. |

## 1. Introduction

One of the most effective ways of communicating information is the speech. Speech includes verbal content as well as gestures, facial expressions, intonation, and emotional content. Emotional state information to be acquired during speech is also a part of communication. During a speech, it is possible to infer a person's state of mind (emotion), gender, attitude, dialect, and so on. Today, with the development of web and mobile applications, human-machine communication has increased. Voice channels are now used in human-machine communication [1]. Increasingly, voice assistance tools are taking a part in answering questions and fulfilling requests instantly and correctly. as they become more common in our daily interactions. For example, Virtual Personal Assistants (VPAs) usage such as Cortona, Apple's Siri, Google Assistant, and Amazon Alexa is increasing [2]. Although these Virtual Personal Assistants understand the necessary commands by inferring words,

they are not good enough at understanding people's emotions and reacting accordingly. Thence, emotion recognition is becoming an investigative field for computer science [3]. People express their emotions using their language. However, in addition to this, qualities such as intonation and speech rate during speech can also be processed and analyzed using signal processing and audio processing [4]. In emotion analysis studies, two approaches are generally applied in the data collection phase. The first one is to attach some sensors to the appropriate parts of the body that give clues about the emotional state. In this approach, data is collected through the interaction of the human body. The other approach is that instead of human-body interaction, human outputs such as sounds or movements are captured using a recording device. The collected data is processed and analyzed for mood [3].

Numerous studies in the literature focus on voice-based sentiment analysis. Some studies on the TESS dataset are as

follows. One of them, Venkataramanan et al. made use of The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset in their 2019 research. After preprocessing the audio signals, they performed experiments by creating a four-layer 2D-CNN architecture. Their experiments resulted in the highest accuracy of 68% [2]. In their 2022 research, Donuk et al. introduced a model by designing an LSTM architecture. To assess its performance, they utilized a dataset that combined both the RAVDESS and TESS datasets. The model demonstrated an impressive 88.92% accuracy in classifying eight distinct emotions [5]. In another study, Akinpelu et al. used the Erling Emotional Speech Database (EMO-DB) dataset and the TESS dataset together. As a result of the experiments, they obtained 97.2% accuracy with TESS data and 94.3% accuracy with the EMO-DB dataset [6]. Patel et al. in their 2020 study using the TESS dataset obtained a remarkable accuracy rate of 96% using an autoencoder [7]. In 2021, a speech-to-emotion recognition method using the TESS and RAVDESS datasets is proposed by Asiya et al. In their study, they implemented a deep learning-based classification model of emotions produced by speech based on voice data such as mel spectrogram, chromatogram, and Mel Frequency Cepstral Coefficient (MFCC). That model achieved 89% accuracy using RAVDESS and TESS datasets and various data augmentation techniques [8]. In their study conducted in 2022, Gokilavani et al. performed sentiment analysis with CREMA-D, RAVDNESS, and TESS datasets. They achieved 96% accuracy for the Ravdness dataset, 99% accuracy for the TESS dataset, and 84% accuracy for the Crema-D dataset in their models using the CNN model [9].

In this study, signal processing methods have been used first to classify emotion from voice signals. Different models have been created with deep learning methods to classify the preprocessed voice signals and the most successful method is proposed as a decision support system.

## 2. Methodology

The TESS (Toronto emotional speech set) dataset has been used to classify emotion from human voice signals [10]. With the proposed model, human voices in the dataset are classified. In this section, in addition to a review of the dataset used, a detailed analysis of MFCC, the method utilized for the extraction features from voice data, and the deep learning models that are developed for the classification of voice data are presented.

### 2.1 Dataset

This study focuses on using the TESS dataset to classify emotions from voice signals. The TESS dataset is considered one of the largest datasets of human speech sounds available and has a balanced mix of male and female voice recordings. This balance is advantageous in the training step since other datasets may contain only one gender, which can lead to

biased results. The TESS dataset contains a total of 2800 voice files collected from people aged between 26 and 64 [11]. The dataset used in this study includes voice recordings of both male and female speakers expressing seven different emotions in Table 1 which provides an overview of the number of files for each emotion and speaker gender. To showcase the voice waveforms utilized in the study, Figure 1 illustrates some examples. The objective of this study is to classify emotions from voice signals accurately. By utilizing the TESS dataset, which is more balanced than other datasets, the study aims to achieve more reliable results. This study has promising results to foster the advancement of emotion recognition systems applicable in various domains, including affective computing, healthcare, and education.

### 2.2 MFCC Feature Extraction

In order to solve a problem with machine learning techniques, we need to have the appropriate attributes. However, we may not always have attributes that we can directly use in the problem we are addressing. In such cases, attributes need to be extracted from the data. "Signal Processing" deals with time series [12]. "Image Processing" is the science that deals with visual data such as photos and videos [13]. "Pattern recognition" is the science that aims to extract features from all kinds of signals, which can include both time series and images. When we examine today's studies, it is seen that feature extraction is performed with many different methods and tools. Optimizing the level of features is critical to achieving the highest accuracy in solving the problem [14].

Table 1. Distribution of voice files in the dataset

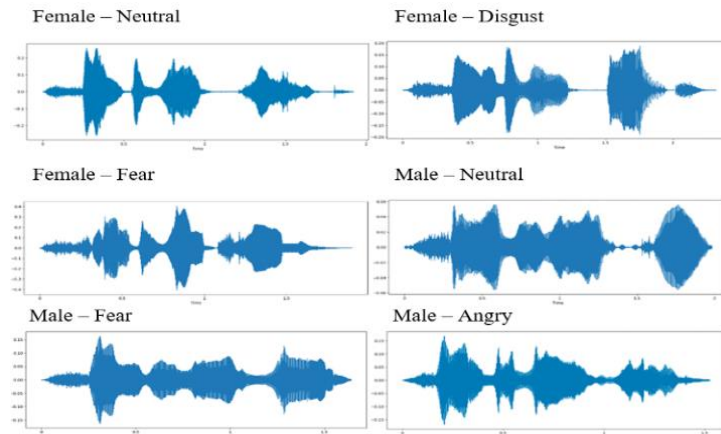| Gender and Emotion | The Number of Sound Recordings |
|---|---|
| Angry - Male | 200 |
| Disgust - Male | 200 |
| Fear - Male | 200 |
| Happy - Male | 200 |
| Neutral - Male | 200 |
| Pleasant Surprise - Male | 200 |
| Sad - Male | 200 |
| Angry - Female | 200 |
| Disgust - Female | 200 |
| Fear - Female | 200 |
| Happy - Female | 200 |
| Neutral - Female | 200 |
| Pleasant Surprise - Female | 200 |
| Sad - Female | 200 |

Figure 1. Representation of human voices as a signal

Utilizing too many attributes can lead to increased costs in the machine-learning process. Therefore, it is essential to carefully select the most relevant features for effectiveness. Selecting the right attributes for the purpose is also very important. In general, two categories of features are used in emotion recognition from speech: prosodic features and vocal tract system features. The first category is derived from prosodic data such as Pitch, Energy, and Duration. The second category is related to the voice path, which includes Cepstrum coefficients such as MFCC, Linear Predicted Cepstrum Coefficients (LPCC), Formants, and Discrete Fourier Transform (DFT) harmonics [15].

This study obtains feature maps of human voices using the MFCC method. MFCC is one of the commonly used techniques in the literature for the extraction of features from audio signals. Before classifying human voices in the TESS dataset, preprocessing is crucial to improve the success rate of machine learning methods. For this reason, the MFCC method is used to extract features from voice recordings. In 1980, Davis and Mermelstein were the first to use the MFCCs method [16]. This method is based on human mimics of the way the human ear perceives sound. The human ear sensitivity is linear up to 1 kHz and linear for higher values continues logarithmically. The transition from the real frequency unit Hertz to the frequency unit Mel is provided by (1) [15].

$$mel(f) = 2595 * \log\left(1 + \frac{f}{700}\right) \qquad (1)$$

The extraction of features from human vocalizations in the TESS dataset was performed utilizing the Librosa library. This library is developed in the Python programming language. This library is used to perform operations on signal recordings [17]. In this study, the Mel-frequency Cepstrum Coefficient (MFCC) method is used to obtain feature maps of human voices. MFCC is a commonly used method in audio signal processing for feature extraction. Before classifying the human voices in the TESS dataset, it is very important to perform preprocessing to increase the success

rate of machine learning methods. For this reason, the MFCC method is used to extract features from the voice recordings. Davis and Mermelstein introduced the MFCC method, which has been found widespread use in speech and voice signal processing, during the 1980s [16]. The MFCC algorithm involves dividing the audio stream into smaller frames using a Hamming window. Spectrums of the frames are then calculated using the Fast Fourier Transform (FFT) and weighted using a Mel scale-based filter bank [18]. Lastly, the Logarithm and Discrete Cosine Transform are applied to calculate the MFCC vector [19]. Figure 2 illustrates the feature processing steps involved in the computation of MFCCs.

### 2.3 Deep Neural Network (DNN) Models

In these models, the model performance can be influenced by the size of the data being used. Handling large datasets can impact the efficiency and effectiveness of the model. Working with small or flat data can lead to overfitting problems for models. To address this challenge and improve the model's performance through effective learning, data augmentation methods are employed. [20]. This can be done by using noise addition, time shifting, and changing speed methods in voice data. In this study, data augmentation is applied using noise addition and time-shifting methods to overcome overfitting and to better train the model. Preprocessing is of great importance for high classification success with machine learning. To this end, the MFCC method has been used to obtain feature maps of voice signals including human voices in the dataset.
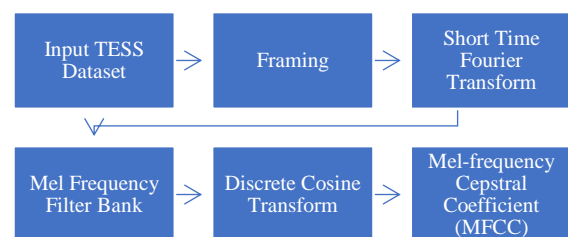


Figure 2. MFCC feature processing steps

Subsequently, these feature maps are fed into the proposed deep models for classification. The classification process has been performed using both CNN and LSTM methods.

The LSTM approach is an extension of the RNN (Recurrent Neural Networks) model and architecture that provides a longer memory span. While RNNs have limited "short-term memory" that utilizes past knowledge in the current neural network, the LSTM method leverages this prior knowledge effectively [21]. In LSTM method artificial neural network structures, the output signal produced in the hidden layer is initialized and is used as one of the values in the next input. The LSTM method is a versatile technique used in several implementations, like speech recognition, anomaly detection in time-series data, handwriting recognition, grammar learning, and music composition [22]. In this study, after obtaining the feature maps from the TESS dataset, the LSTM method is used for classification. The developed LSTM model consists of Dense, Dropout, and Flatten layers and is shown in Table 2. The ReLU layer is used for activation on the obtained feature maps, and the Dropout layer is used to prevent overfitting. Softmax is the preferred choice for classification in the LSTM method. The architectural representation of the LSTM model for emotion classification on speech sounds is illustrated in Figure 3.

The LSTM model is composed of 1 LSTM, 2 Dropout, 1 Flatten, 2 Dense, and 1 Softmax layers. Table 2 shows the parameter, layer, and output shape information of the LSTM model. In deep learning models, the term output shape refers to the ability to efficiently process data sets, which often have variable dimensions. When specifying the dimensions of the output shape, the term 'None' implies that it is flexible depending on the dimensions of the input data.
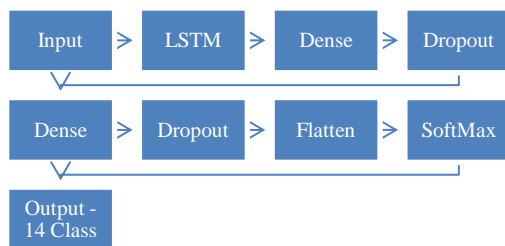


Figure 3. Architectural representation of the proposed LSTM Model

Table 2. Layers and parameters in LSTM model

| Layers | Output shape | Activation function | Parameter |
|---|---|---|---|
| LSTM | (None,123) | - | 61500 |
| Dense | (None, 64) | Relu | 7936 |
| Dropout | (None, 64) | - | 0 |
| Dense | (None, 32) | Relu | 2080 |
| Dropout | (None, 32) | - | 0 |
| Flatten | (None, 32) | - | 0 |
| Dense | (None, 14) | - | 462 |
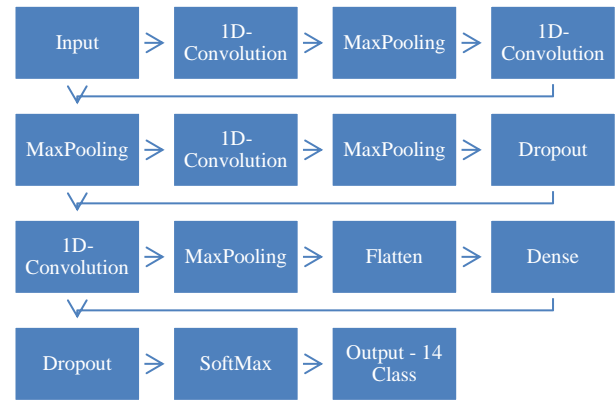


Figure 4. Architectural representation of the proposed CNN Model

One method used for classification is CNN. CNNs have gained widespread adoption in various domains, including image processing, text processing, and audio processing. Their versatile features make them applicable regardless of the number of dimensions [23]. The utilization of 2D-CNN and 3D-CNN has been predominantly applied for image classification purposes, whereas one-dimensional CNNs are commonly employed in various solutions such as signal analysis [24]. In the proposed deep network model for emotion classification from human voices in the TESS dataset, 1D-CNN is used. The proposed model comprises of a 1D-CNN layer, 1D-MaxPooling layer, Dropout layer, Flatten layer, and Dense layer with Softmax activation. The resulting feature maps activate with Relu, and the computational complexity decreases through the utilization of the Pooling layer. During the training process, the Dropout layer has been employed to disable certain nodes in the network to prevent overfitting. The Flatten layer has been responsible for converting the data from a matrix format to a flattened format. The classification process has been accomplished through the utilization of the Softmax layer, which has generated probability values based on the input values. Whichever class is closer to the values obtained, the classification process is completed by placing the data in the relevant class. The CNN model architecture created for emotion classification on speech sounds is given in Figure 4.

When the CNN model is analyzed, it is seen that 4 1D-Convolution, 4 Maxpooling, 1 Dropout, 1 Flatten, 2 Dense, and 1 Softmax layers are implemented. The CNN model's parameters and layer information are presented in detail in Table 3.

## 3. Experimental Result

The Intel DevCloud cloud system has been utilized for preprocessing the voice signals, training the DNN model, and testing the models in this study. To expedite DNN model training and testing, the researchers utilized the Intel oneAPI framework.

Table 3. Layers and parameters in CNN model

| Layers | Output shape | Activation function | Parameter |
|---|---|---|---|
| Conv1d | (None,162,256) | Relu | 1536 |
| Max_pooling1d | (None,81,256) | - | 0 |
| Conv1d | (None,81,256) | Relu | 327936 |
| Max_pooling1d | (None,41,256) | - | 0 |
| Conv1d | (None,41,128) | Relu | 163968 |
| Max_pooling1d | (None,21,128) | - | 0 |
| Dropout | (None,21,128) | - | 0 |
| Conv1d | (None,21,64) | Relu | 41024 |
| Max_pooling1d | (None,11,64) | - | 0 |
| Flatten | (None,704) | - | 0 |
| Dense | (None,32) | Relu | 22560 |
| Dropout | (None,32) | - | 0 |
| Dense | (None,14) | - | 462 |

Furthermore, it should be noted that oneAPI is an open, cross-industry, standards-based programming model that supports multiple architectures and vendors. By providing a unified development experience across accelerator architectures, oneAPI aims to enhance application performance, boost productivity, and encourage innovation. The oneAPI platform ensures the improved code to profit by various hardware architectures, including GPUs, multi-core CPUs, or other hardware, all through singular sources [25]. This shortened the experiment times considerably. In the experiments, the TESS dataset is used for the training and testing stages. Different neural network models are created to perform emotion classification from voice signals using different architectures. Accomplishment evaluation of the models is compared on Accuracy, Recall (Sensitivity), Precision,and F-score.

*Accuracy*

The accuracy of a model is computed by the ratio of the number of correct predictions to the total number of predictions made on the entire dataset(2).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (2)$$

The symbols TP, TN, FP, and FN state true positive, true negative, false positive, and false negative values, respectively.

*Recall*

Recall, called sensitivity or true positive rate, is a performance metric that represents the number of true positive predictions made by the model divided by the total number of actual positive instances in the dataset (3).

$$Recall = \frac{TP}{FN} \qquad (3)$$

*Precision*

Precision is a performance metric as given by (4).

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

*F Score*

The F score is used to show the trade-off between sensitivity and recall. The F score is obtained by equation (5).

$$F\ Score = 2 \times \frac{precision \times recall}{precision+recall} \qquad (5)$$

Equations (2), (3), (4), and (5) are utilized to compare the performance of different models. In classification problems, the prediction results can be summarized using a confusion matrix, which is a tabular representation. As shown in Table 4, the confusion matrix can be used to get an idea of the errors made by the classifier. The confusion matrix also gives insight into the error types made by the model.

The study conducted emotion classification by first obtaining feature maps through the MFCC method. The feature maps are then used as input to deep neural network models for classification. 90% and 10% of the dataset are used for training and for testing the models, respectively. The study employed two popular deep learning models, namely CNN and LSTM, for the classification task. The LSTM method, which is widely used in various applications such as sentiment analysis, text generation, and time series, proved to be highly accurate in sentiment analysis from voice. Table 5 provides a comprehensive summary of the LSTM-based model's performance, including precision, recall, and F score. Moreover, the study presented the confusion matrix of the classification test in Figure 5, demonstrating the model's ability to correctly classify samples into their respective emotion classes. Furthermore, the results suggest that the LSTM-based model has successfully achieved accurate emotion classification from voice signals, demonstrating its potential for practical applications in affective computing, healthcare, and education. The precision, recall, and F score provided in Table 5 indicate the model's strong performance in recognizing male and female voices expressing various emotions. Overall, the study's findings provide valuable insights into the effectiveness of deep learning methods, particularly LSTM, for emotion classification from voice signals

When Table 5 is examined for the model created with the LSTM method, it is shown that the most successful classes are Male-Neutral, Male-Sad, Female-Happy, and Female-Sad. The most unsuccessful class is Male - Happy.

Table 4. Confusion Matrix

| | Positive Prediction | Negative Prediction |
|---|---|---|
| **Actual Positive** | True Positive (TP) | False Negative (FN) |
| **Actual Negative** | False Positive (FP) | True Negative (TN) |

.

Table 5. Classification Evaluations Obtained In The LSTM Model

| Classes | Recall | Precision | F score |
|---|---|---|---|
| Angry - Male | 0.87 | 0.96 | 0.91 |
| Disgust - Male | 0.74 | 0.81 | 0.77 |
| Fear - Male | 0.98 | 0.91 | 0.95 |
| Happy - Male | 0.65 | 0.58 | 0.61 |
| Neutral - Male | 0.98 | 0.98 | 0.98 |
| Pleasant Surprise - Male | 0.77 | 0.72 | 0.74 |
| Sad - Male | 0.97 | 1.00 | 0.98 |
| Angry - Female | 0.90 | 0.98 | 0.94 |
| Disgust - Female | 0.95 | 0.92 | 0.94 |
| Fear - Female | 0.97 | 0.92 | 0.94 |
| Happy - Female | 0.97 | 1.00 | 0.98 |
| Neutral - Female | 0.91 | 1.00 | 0.95 |
| Pleasant Surprise - Female | 0.96 | 0.91 | 0.94 |
| Sad - Female | 0.97 | 0.98 | 0.98 |

The average accuracy for all classes for the LSTM model is 0.90. Figure 6 illustrates the accuracy and loss curves of the model created using the LSTM method.

When Figure 6 is examined, it becomes evident that the accuracy of the LSTM model is above 90% on the training data, indicating its success during the training phase. However, when examining the test data, the decrease in accuracy suggests that the model is less specialized for the test data, and its generalization ability is limited. Similarly, the loss value in the loss curve increases as it approaches the training data, indicating that the model is inclined to make more errors when applied to the test data.

The effectiveness of the CNN method in performing emotion analysis on voice signals underwent evaluation.
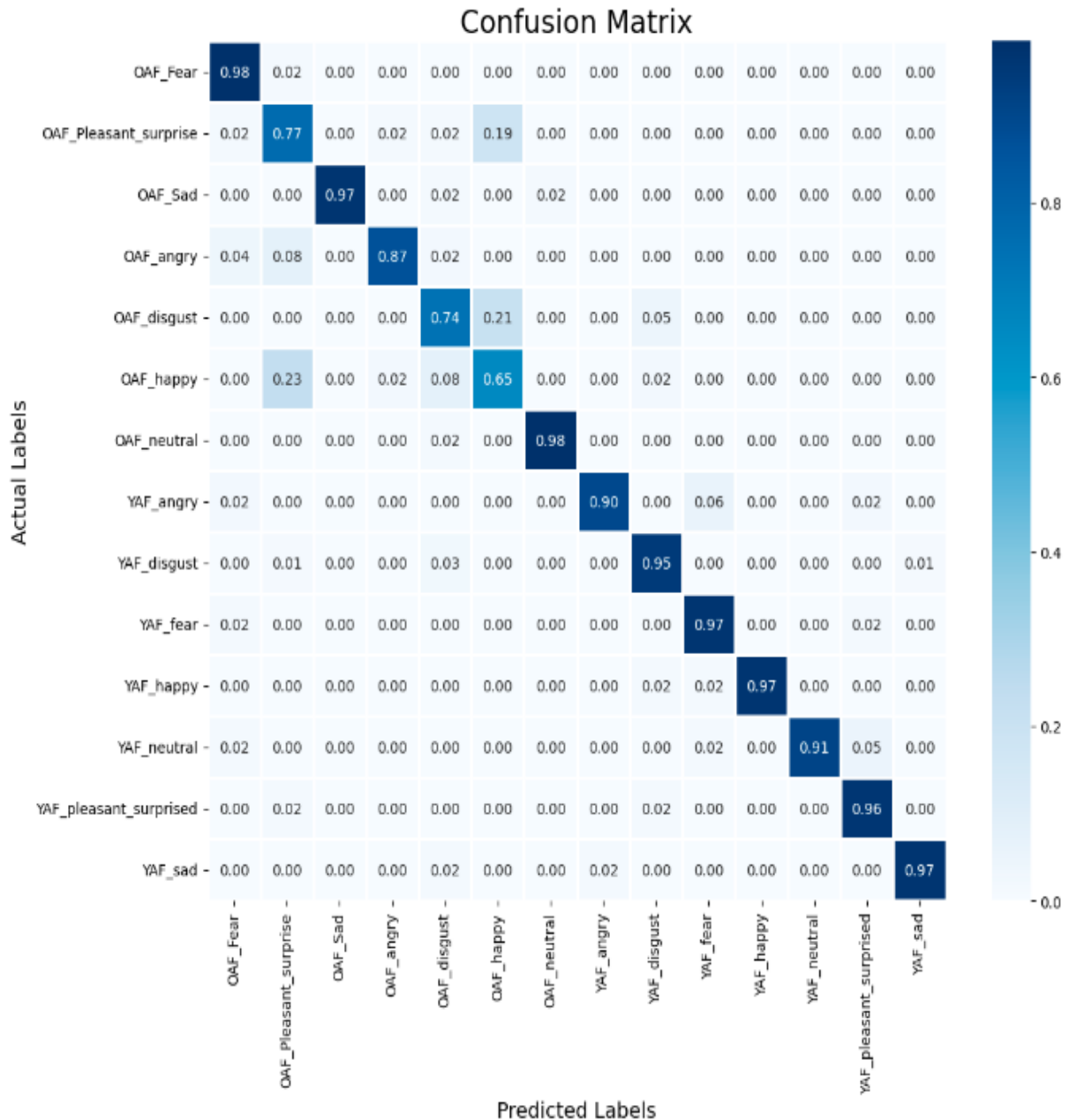


Figure 5. Confusion matrix obtained in the LSTM model.
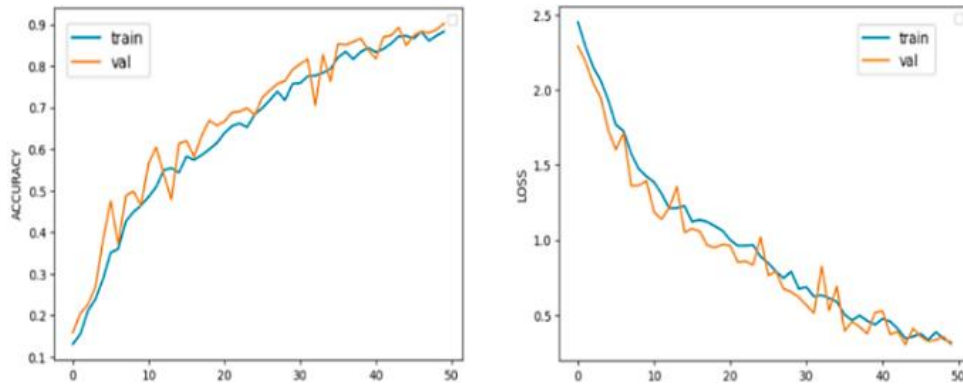
Figure 6. Accuracy and loss curves for the LSTM model
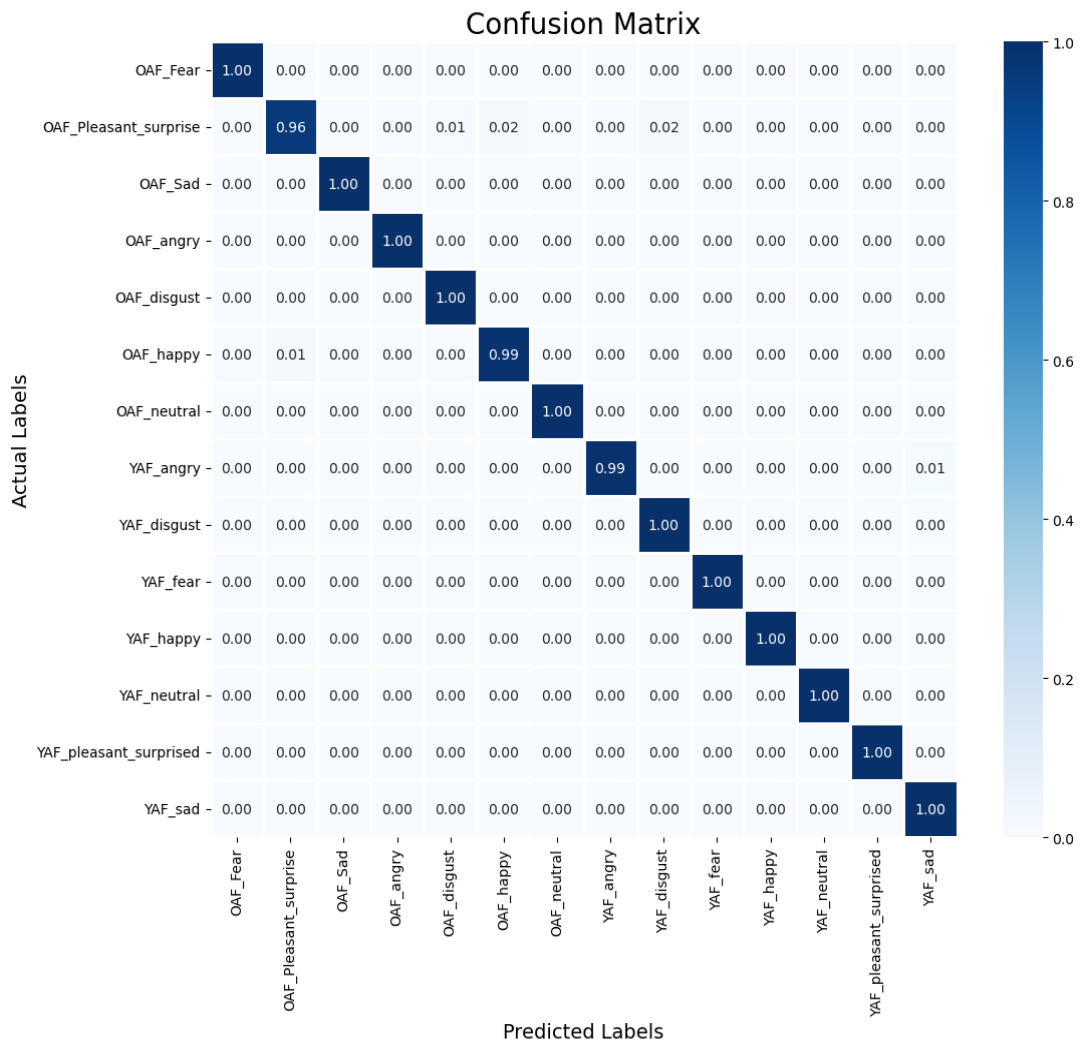
## Confusion Matrix



Figure 7. Confusion matrix obtained in the CNN model

Following the creation of a model using the CNN method, testing is conducted, and the resulting confusion matrix is presented in Figure 7.

Moreover, the accuracy, sensitivity, and F Score evaluations of the CNN method, which achieved high accuracy rates, are given in Table 6.

When the test results of the CNN model are analyzed in Table 6 and the confusion matrix is analyzed in Figure 7, it is seen that 10 out of 14 classes achieved full success. The remaining 4 classes achieved results close to full success.

In the deep neural network model created using the CNN method, the average accuracy value in all classes is 0.995. Figure 8 illustrates the accuracy and loss curves of the CNN method, which exhibited superior classification performance compared to the LSTM method.

In the training phase, as illustrated in Figure 8, the CNN model exhibits an accuracy of over 99% on the training data. This indicates superior performance when compared to the LSTM model.

Table 6. Classification Evaluations Obtained In The CNN Model

| Classes | Recall | Precision | F score |
|---|---|---|---|
| Angry - Male | 0.99 | 1.00 | 1.00 |
| Disgust - Male | 0.99 | 0.96 | 0.97 |
| Fear - Male | 1.00 | 1.00 | 1.00 |
| Happy - Male | 0.96 | 0.99 | 0.98 |
| Neutral - Male | 1.00 | 1.00 | 1.00 |
| Pleasant Surprise - Male | 0.96 | 0.93 | 0.95 |
| Sad - Male | 0.99 | 1.00 | 1.00 |
| Angry - Female | 0.99 | 1.00 | 1.00 |
| Disgust - Female | 0.99 | 0.99 | 0.99 |
| Fear - Female | 1.00 | 1.00 | 1.00 |
| Happy - Female | 1.00 | 1.00 | 1.00 |
| Neutral - Female | 1.00 | 1.00 | 1.00 |
| Pleasant Surprise - Female | 0.99 | 1.00 | 1.00 |
| Sad - Female | 1.00 | 1.00 | 1.00 |

Furthermore, the consistent decrease in the loss value during training suggests that the CNN model effectively mitigates overfitting. Lastly, the accuracy value remains relatively stable when applied to the test data, indicating the model's strong generalization ability. These results underscore the CNN model's superior efficacy compared to the LSTM model.

To assess the effectiveness of the our approach, we compare the performance with similar studies conducted on the TESS dataset in the existing literature. A comparative analysis of studies providing deep learning models for speech emotion recognition is presented in Table 7.

It is evident that our model outperforms most of the previous studies on the same datasets. In deep network architectures, the preprocessing and the amount of data used during training are very important for accuracy. In this study, feature extraction took place using the MFCC method. In addition, data augmentation operations are applied to diversify the dataset. This helps to reduce overfitting and increase the generalization ability of the model. In addition to these factors, the architecture of our CNN model is also effective in its high success. The model encloses three convolutional layers, two max-pooling layers, two fully connected layers, and a dropout layer. The local features are extracted from the voice data by convolutional layers, while the size of the feature maps is decreased by the max pooling layers. The other layers, fully connected layers, enable the classification of speech into predefined categories by learning complex features from the extracted data. For these reasons, our results are more successful than previous studies despite the larger number of classes. The CNN model performance is remarkable as it achieves error-free classification results in 10 out of 14 emotion and gender classes. In particular, in male classification, the model showed error-free classification in various emotional states and achieved F1 scores of 1.00 in categories such as neutral, fear, angry, and sad.
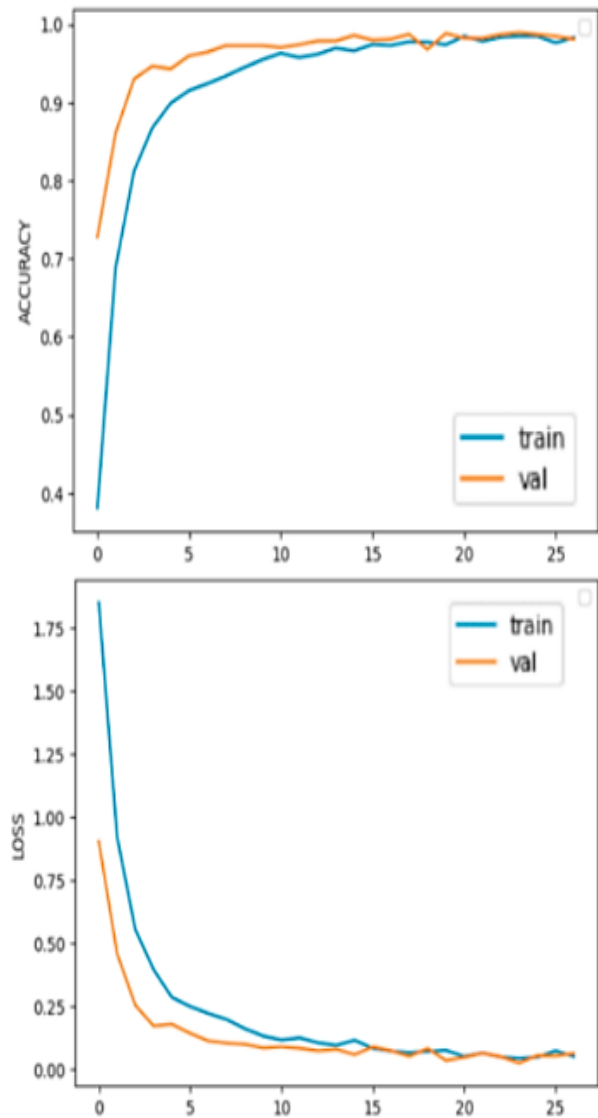


Figure 8. Accuracy and loss curves for the CNN model

Similarly, in the female classification, the model demonstrated high accuracy with error-free classification in emotions such as neutral, angry, happy, fear, pleasant surprise, and sad, recording an F1 score of 1.00. This accuracy value demonstrates the robustness and adaptability of the model in discriminating complex emotional nuances and gender categories across a wide range of classes.

Table 7. Comparative study on TESS dataset

| Reference | Method | No. of Class | Accuracy (%) |
|---|---|---|---|
| Venkataramanan et al. [2] | CNN | 14 | 68 |
| Donuk et al. [5] | LSTM | 8 | 88.92 |
| Akinpelu et al. [6] | MLP&SVM | 6 | 97.20 |
| Patel et al. [7] | CNN | 7 | 96 |
| Asiya et al. [8] | CNN | 8 | 89 |
| Gokilayani et al. [9] | CNN | 7 | 99 |
| *Proposed method* | **A new CNN model** | *14* | *99.5* |

## 4. Conclusions

Sentiment analysis has gained immense popularity in recent years with companies, organizations, and governments using it to understand people's reactions to various topics. Sentiment analysis can be carried out for different types of data, comprising textual and auditory content. In this study, sentiment classification is conducted on voice signals using the TESS dataset. This dataset consists of 14 classes, including 7 emotions and gender classification (male-female). Short-Time Fourier Transform and MFCC methods are used for feature extraction. Diversity in the dataset is ensured by data augmentation. This helps to lessen overlearning and enhance the model generalization ability. After data augmentation, deep neural network models are trained for classification. While 90% of the dataset was used for training, 10% was used for testing. The performances of two deep learning models, namely CNN and LSTM, are compared. The LSTM model achieved an accuracy of 90% with 14 classifications, while the CNN model outperformed with a 99.5% accuracy rate, achieving complete success in 10 out of 14 classes, with near-complete success in the remaining 4 classes. The proposed CNN model aims to classify gender and emotion from human voices in the TESS dataset. The model's architecture embodies three convolutional layers, two max-pooling layers, two fully connected layers, and a dropout layer. The convolutional layers extract increasingly complex features from the data, while the max pooling layers help to abates the dimension of the data and improve the model's generalization ability. The fully connected layers learn a complex representation of the data and map it to the output classes. The dropout layer assures a prevention of overfitting by randomly dropping out some of the neurons during training. Comparing the success rate of the proposed model with other similar works in the literature yields that the proposed model achieves a higher accuracy rate, which is one of the highest rates reported in recent studies using the TESS dataset. Therefore, the results of proposed method demonstrate the effectiveness of deep learning methods, especially the CNN model, in achieving high accuracy rates in sentiment analysis using voice data.

## Declaration

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

## Author Contributions

Fatih Şengül conceptualized and conducted the study, authored the manuscript, and actively contributed to the entire research process. Sıtkı Akkaya provided valuable guidance and supervisory support throughout the study, offering insightful feedback and contributing to the refinement of the research methodology. Both authors collaboratively wrote and reviewed the manuscript. The authors are listed in alphabetical order.

## Acknowledgment

## Nomenclature

*MFCC*  : Mel Frequency Cepstral Coefficient
*LSTM*  : Long Short-Term Memory
*CNN*  : Convolutional Neural Network
TESS  : Toronto emotional speech set

## References

1. Liu, K., Wang, D., Wu, D., Liu, Y., and Feng, J., *Speech emotion recognition via multi-level attention network.* IEEE Signal Processing Letters, 2022. **29**: p. 2278-2282.

2. Venkataramanan, K., and Rajamohan, H. R., *Emotion recognition from speech.* arXiv preprint, 2019. p. 1912-10458.

3. Aydin, M., Tuğrul, B., and Yilmaz, A. R., *Emotion Recognition System from Speech using Convolutional Neural Networks.* Computer Science, 2022. p. 137-143.

4. Xu, Y., *English speech recognition and evaluation of pronunciation quality using deep learning.* Mobile Information Systems, 2022. p. 1-12.

5. Donuk, K., and Hanbay, D., *Konuşma Duygu Tanıma için Akustik Özelliklere Dayalı LSTM Tabanlı Bir Yaklaşım.* Computer Science, 2022. **7**(2): p. 54-67.

6. Akinpelu, S., and Viriri, S., *Robust Feature Selection-Based Speech Emotion Classification Using Deep Transfer Learning.* Applied Sciences, 2022. **12**(16): 8265.

7. Patel, N., Patel, S., and Mankad, S. H., *Impact of autoencoder based compact representation on emotion detection from audio.* Journal of Ambient Intelligence and Humanized Computing, 2022. p. 1-19.

8. Asiya, U. A., and Kiran, V. K., *Speech Emotion Recognition-A Deep Learning Approach*, in Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). 2021. Palladam, India: p. 867-871.

9. Gokilavani, M., Katakam, H., Basheer, S. A., and Srinivas, P. V. V. S., *Ravdness, crema-d, tess based algorithm for emotion recognition using speech,* in 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT). 2022. Tirunelveli, India: p. 1625-1631.

10. Pichora-Fuller, M. K., and Dupuis, K., *Toronto emotional speech set (TESS). Scholars Portal Dataverse*. 2020. *1*.

11. Sun, C., Li, H., and Ma, L., *Speech emotion recognition based on improved masking EMD and convolutional recurrent neural network*. Frontiers in Psychology, 2023. **13**: 1075624.

12. Zhang, C., Mousavi, A. A., Masri, S. F., Gholipour, G., Yan, K., and Li, X., *Vibration feature extraction using signal processing techniques for structural health monitoring: A review*. Mechanical Systems and Signal Processing, 2022. **177**: 109175.

13. Zhang, Y., and Zheng, X., *Development of Image Processing Based on Deep Learning Algorithm*, in 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). 2022. Dalian, China: p. 1226-1228.

14. Jastrzebska, A., *Time series classification through visual pattern recognition*. Journal of King Saud University-Computer and Information Sciences, 2022. **34**(2): p. 134-142.

15. Kop, B. Ş., and Bayindir, L., *Bebek Ağlamalarının Makine Öğrenmesi Yöntemleriyle Sınıflandırılması*. Avrupa Bilim ve Teknoloji Dergisi, 2021. **27**: p. 784-791.

16. Davis, S., and Mermelstein, P., *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*. IEEE transactions on acoustics, speech, and signal processing, 1980. **28**(4): p. 357-366.

17. Choudhary, R. R., Meena, G., and Mohbey, K. K., *Speech emotion-based sentiment recognition using deep neural networks*, in Journal of Physics: Conference Series. IOP Publishing. 2022. p. 012003.

18. Yıldırım, M., *MFCC Yöntemi ve Önerilen Derin Model ile Çevresel Seslerin Otomatik Olarak Sınıflandırılması*. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 2022. **34**(1): p. 449-457.

19. Li, F., Liu, M., Zhao, Y., Kong, L., Dong, L., Liu, X., and Hui, M., *Feature extraction and classification of heart sound using 1D convolutional neural networks*. Eurasip Journal on Advances in Signal Processing, 2019. **2019**(1): p. 1-11.

20. Maharana, K., Mondal, S., and Nemade, B., *A review: Data pre-processing and data augmentation techniques*. Global Transitions Proceedings, 2022. **3**(1): p. 91-99.

21. Alpay, Ö., *LSTM mimarisi kullanarak USD/TRY fiyat tahmini*. Avrupa Bilim ve Teknoloji Dergisi, 2020. p. 452-456.

22. Priyadarshini, I., and Puri, V., *Mars weather data analysis using machine learning techniques*. Earth Science Informatics, 2021. **14**: p. 1885-1898.

23. Adem, K. and Kılıçarslan, S., *COVID-19 diagnosis prediction in emergency care patients using convolutional neural network*. Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 2021. **21**(2): p. 300-309.

24. Liu, S., and Chen, M., *Wire Rope Defect Recognition Method Based on MFL Signal Analysis and 1D-CNNs*. Sensors, 2023. **23**(7): p. 3366.

25. Christgau, S., and Steinke, T., *Porting a legacy cuda stencil code to oneapi*, in 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2020. New Orleans, LA, USA: p. 359-367.

*Research Article*

# Ecological dyeing and UV-protective functionalization of cotton/lyocell blend fabrics designed for high comfort summer clothing

*Nazli Uren* [a,*] (iD)

[a]*Department of Textile Engineering, Dokuz Eylul University, Izmir 35390, Turkey*

| ARTICLE INFO | ABSTRACT |
|---|---|
| | Wearing clothes that absorb or block harmful UV radiation is one of the most effective forms of protection against sun damage and skin cancer. In the current study, sustainable processes which will provide high clothing comfort and a sufficient level of UV protection for fabrics used in production of lightweight sun-protective summer clothing were proposed. In accordance with the scope of the study, structural properties, low-stress mechanical properties, surface characteristics, permeability, and UV-protection properties of three woven fabrics produced with different weft settings and lyocell ratios were determined. The fabric type which had the most desirable results was selected for dyeing experiments. Avocado (*Persea americana*) seed was used as an eco-friendly source for dyeing and UV-protective functionalization of cotton/lyocell blend fabric. To obtain different hues, the fabrics were mordanted according to two different recipes using magnesium sulphate and ferrous sulphate and dyed with the natural colorant extracted from avocado seeds. Comfort, color, fastness, and UV-protective properties of dyed samples were evaluated by laboratory tests. Results indicated that it is possible to achieve an excellent level of UV protection (UPF 50+) by dyeing cotton/lyocell blend fabrics with avocado seed extract and produce summer clothing with improved comfort and UV-protective properties, without using any toxic materials. |

## 1. Introduction

Ultraviolet (UV) radiation is a form of electromagnetic radiation that comes from the sun and several man-made sources. The most common types of skin cancer are found on sun-exposed parts of the body, and their occurrence is commonly related to exposure to the UV rays in sunlight. Melanoma is a more serious but less common type of skin cancer, and it is also related to sun exposure [1].

UV radiation is divided into three groups: UVA, UVB and UVC. Exposure to UVA rays can lead to aging of skin cells and cause some indirect damage to cells' DNA. UVB rays are the main cause of sunburn, and they can also damage the DNA in skin cells [1]. UVC reacts with ozone in the atmosphere, consequently, it doesn't reach the ground.

Recent studies indicate that insufficient sun exposure may be responsible for increased incidences of several health problems such as breast cancer, colorectal cancer, hypertension, cardiovascular disease, metabolic syndrome, multiple sclerosis, Alzheimer's disease, autism, asthma, type 1 diabetes, and myopia [2]. Recognition of the beneficial effects of UV exposure has led to a reconsideration of sun avoidance policies [3,4]. As it provides several health benefits, it is not suggested to avoid sunlight completely, but it is necessary to limit the duration and intensity of the exposure. Staying in the shade during midday hours, protecting the skin with proper clothing and using sunscreens are common ways for UV protection.

Thermo-physiological comfort is a very important aspect to be considered, especially for summer clothing [5-9]. Lightweight fabrics made of cotton and regenerated cellulosic fibers such as viscose, modal and lyocell are often preferred in summer clothing because of their good tactile comfort, high air permeability and high moisture vapor transmission rates [10]. However, the level of UV protection provided by textiles greatly changes depending on several parameters such as raw material, fabric construction, color and application of functional finishes [10,11]. Lightweight cotton fabrics with bright colors and low coverages may not provide the required level of protection against UV rays [12]. Researchers previously reported that textiles dyed with natural colorants may

exhibit increased levels of UV protection [13-21]. Natural dyes are renewable and sustainable bioresource products. In this respect, dyeing fabrics with natural colorants can be a sustainable option for producing sun-protective summer clothing. Using natural sources can also minimize the several negative effects of synthetic dyes [20-28].

Another important aspect of sustainable production is the fiber content. Cotton is a natural fiber which is commonly preferred in summer clothing. However, producing 1 kg of cotton fiber requires more than 20000 L of water [29]. Moreover, cotton is responsible for 11% and 24% of the world's pesticides and insecticides consumption, respectively. Lyocell is a biodegradable regenerated cellulosic fiber which requires a lot less dye than cotton and no bleaching, thus it is a more sustainable option in terms of energy and water consumption [30]. Besides eco-sustainability, using lyocell and its blends in garments provides a series of desirable attributes such as comfort, efficient moisture transportation, vivid colors with high fastness grades, an elegant drape, a strong durability, and antibacterial properties [29-32].

There are several studies investigating the use of natural dyes for sustainable coloration of lyocell fabrics [20-26]. A review of the previous literature indicated that UV-protective properties of lyocell/cotton blend fabrics can be successfully enhanced using aqueous extracts of several plants including kermes oak (*Quercus coccifera* L.) seeds and mango (*Mangifera indica* L.) bio-wastes (seed and peel) [20,21].

The estimated total world production for avocados (*Persea americana*) in 2022 was over 8.9 million tons [33]. In this respect, bio-waste of avocado fruit - which is approximately 27% the of the total mass - can be a favorable source for coloration of textiles. Even though it was previously reported that cotton and wool fabrics can be successfully dyed with pigments extracted from avocado seeds and peels [34-36], coloration and UV-protective functionalization of lyocell blended fabrics using avocado bio-wastes was not discussed. In a study carried out by Cuk and Gorjanc, cotton was dyed with extracts of curcuma, green tea, avocado seed, pomegranate peel and horse chestnut bark [36]. However, the reported UV protection levels achieved by applying avocado seed extract on cotton fabric by the suggested extraction and dyeing methods were quite low.

The aim of the current study was to propose a fabric design and sustainable processes which will provide the desired qualities such as high clothing comfort and a sufficient level of UV protection for producing lightweight sun-protective summer clothing. In accordance with this purpose, several components of clothing comfort were investigated by laboratory tests for three fabrics made of cotton and lyocell fibers. The fabric which provided the most desirable results were dyed with avocado seed extract according to three different recipes. The dyeability of cotton and lyocell fibers with the proposed natural colorant was investigated based on color and fastness results. The efficiency of avocado seed extract as a natural agent for UV-protective functionalization was also discussed.

## 2. Materials and Methods

### 2.1 Material

The fabrics used in the current study have a 5-end warp-dominant satin weave pattern, produced with identical yarns and weave parameters (except weft setting). Yarn structure has a large effect on fabric comfort. For instance, woven fabrics produced with compact-spun yarns may possess better air and water vapor permeabilities when compared to fabrics produced with ring-spun yarns [37]. For the fabrics investigated in the current study (LL, ML, HL), compact yarns were used in both warp and weft directions. The warp warns were made of 100% combed cotton and the weft yarns were made of 100% lyocell fibers (Table 1).

Aqueous extract of avocado (*Persea americana*, Hass cultivar) seed was used for dyeing experiments. Fully ripe avocados were procured from a local market. $Mg(SO_4) \cdot 7 H_2O$ (magnesium sulfate), $FeSO_4 \cdot 7 H_2O$ (ferrous sulfate), and $NaHCO_3$ were supplied from Merck (Darmstadt, Germany).

### 2.2 Extraction

The seeds of avocado fruit were separated from the pulp and rested at room temperature for two hours. The thin outer shell of the seeds was peeled, and the seeds were dipped into the 0.01M $NaHCO_3$ solution with a material to liquor ratio of 1:20.

The extraction was carried out in a beaker at 85 °C and the duration of extraction was three hours. During the extraction, the liquor volume was kept constant by covering the top of the beaker with an elastic cover. When the extraction was finalized, solid residues of avocado seeds were separated from the dye extract by using a 100% cotton filtering fabric.

### 2.3 Dyeing

For dyeing, samples ML1, ML2, and ML3 were dipped into the prepared avocado seed extract with an M:L ratio of 1:20. Dyeing was carried out by resting the samples in the prepared dye solutions for 3 hours, at room temperature, inside glass bottles with sealed caps. During the 3-hour dyeing process, for each 20 minutes of resting, the dye solution and the sample were mixed by manually shaking the bottles for 5 seconds. After 3 hours of dyeing, samples were rinsed with running tap water for five minutes and rested on a hanger for drying.

Table 1. Production parameters of the studied fabrics

| Sample code | Weave pattern | Yarn properties | | Setting (cm$^{-1}$) | |
|---|---|---|---|---|---|
| | | *Warp* | *Weft* | *Warp* | *Weft* |
| LL | 5-end satin | Cotton (9.8 tex) | Lyocell (9.8 tex) | 82 | 40 |
| ML | 5-end satin | Cotton (9.8 tex) | Lyocell (9.8 tex) | 82 | 45 |
| HL | 5-end satin | Cotton (9.8 tex) | Lyocell (9.8 tex) | 82 | 50 |

## 2.4 Mordanting

To obtain different hues, the fabrics were mordanted with nontoxic metallic salts - magnesium sulphate and ferrous sulphate. Sample ML2 was pre-mordanted with magnesium sulphate and simultaneously mordanted with ferrous sulphate. Sample ML3 was only pre-mordanted with ferrous sulphate. No mordanting process was carried out for sample ML1.

- Pre-mordanting: For pre-mordanting, two separate solutions were prepared. Magnesium sulphate (5% owf) and ferrous sulphate (2% owf) were dissolved in distilled water and the samples (ML2 and ML3 respectively) were rested in prepared mordant solutions at room temperature for 24 hours. No rinsing was carried out after pre-mordanting. Mordanted samples were dried on a hanger.

- Simultaneous mordanting: Simultaneous mordanting was applied to sample ML2 only. For simultaneous mordanting, sample ML2 was removed from the dye bath at the end of the first hour of dyeing. When the fabric sample was removed, 0.35% owf of ferrous sulphate was dissolved in the dye bath, and sample ML2 was dipped back into the mordant added dye solution. The sample was then rested in the mordant added dye solution for two more hours. The total dyeing duration of sample ML2 was three hours, which was equal to the dyeing duration of sample ML1 and ML3.

## 2.5 Measurements

Structural parameters (mass per unit area and thickness), low-stress mechanical properties (extensibility and bending rigidity), out-of-plane deformation and recovery behavior (deformation, elasticity, plasticity, and hysteresis), surface characteristics (micro and macro-surface variations), tactile comfort, permeability (air and water vapor), color, fastness and UV-protective properties of fabrics were determined by laboratory tests. The samples were conditioned before measurements as prescribed in TS EN ISO 139/A1. All measurements were carried out at standard atmospheric conditions.

- Structural properties: Mass per unit area was measured according to TS EN 12127. The thickness of fabrics was measured under 5 gf/cm$^2$ pressure using James Heal R&B Cloth Thickness Tester having a circular presser foot with a size of 100 mm$^2$. Five trials were done for each fabric type.

- Extensibility: Extensibilities in warp and weft directions were determined using Instron 4411 Universal Tensile Tester. The sample size was 250 mm x 50 mm, the test length was 150 mm, the test speed was 25 mm/min, and the pre-tension was equal to zero. Percent elongation values recorded under 100 N/m load were presented as extensibility. Five trials were done for each test direction and fabric type.

- Bending rigidity: Bending rigidity was determined as prescribed in ASTM D1388-18 Option A. Bending properties were determined for warp, weft and two diagonal (-45° and +45°) directions. Five trials were done

for each test direction and fabric type.

- Out-of-plane deformation: Deformation and recovery behavior of fabrics in out-of-plane deformation state was investigated using a multifunctional test instrument "Tactile Sensation Analyzer" (TSA) developed by the German company emtec Electronic GmbH [38,39]. For the measurements, the sample (120 mm x120 mm) was fixed to the circular frame and deformed in the out-of-plane direction by the movement of the measuring head of the device. When a load of 600 mN is reached, the magnitude of deformation (D) was recorded. Then the same sample was automatically deformed by the test device for the second time, and the magnitude of deformation observed in this cycle was recorded as elasticity (E). The energy of recovery after the first cycle was recorded as hysteresis (H) and the magnitude of permanent deformation was recorded as plasticity (P) [38]. Five samples were tested for each fabric type.

- Surface characteristics: Macro-surface variations and micro-surface variations of face and back sides of the fabrics were measured by Tactile Sensation Analyzer. For investigating surface characteristics, the test sample was fixed to the circular frame. When the measuring head contacted with the sample, applying a 100 mN load, the head rotated on the fabric surface, and the sound intensity peaks at two frequency regions were recorded using a special sound analysis technique. These two peaks were named as TS7 and TS750 which refer to micro-surface variations and macro-surface variations, respectively. For the measurements, five samples (120 mm x 120 mm) were tested for each sides of the fabrics.

- Nozzle test: Nozzle test is a conventional test method where the force required to pull a fabric sample through a ring, or a nozzle is determined [40-42]. It was previously reported that the pulling forces recorded during nozzle test were significantly correlated with tactile comfort scores of fabrics determined by sensory evaluations [41,42]. In the current study, a nozzle construction with 24 mm diameter and 20 mm height was used. For the measurements, a circular sample with 100 mm diameter was attached to a needle, pulled through the nozzle with a speed of 25 mm/min and the maximum pulling force was recorded. For each fabric type, five trials were done.

- Air permeability: Air permeability was determined according to TS 391 EN ISO 9237 using Textest FX 3300 Air Permeability Tester III. The test area was selected as 20 cm$^2$ and the test pressure was 100 Pa. 10 measurements were done for each fabric type.

- Water vapor permeability: Water vapor permeability was measured as prescribed in BS 7209:1990 using SDL ATLAS equipment no M261. The duration of the measurements was 24 hours. Three trials were done for each fabric type. Permeability was tested with the principle of measuring the water vapor movement from back side towards the face side.

- Color and fastness properties: Color of samples were measured using Minolta Spectrophotometer CM-3600d (illuminant D65 and 10° observer angle) and presented in

CIE L* a* b*, C*, h° coordinate system. Color strength (K/S) of dyed samples were given as well. Fastness to washing was measured according to ISO 105-C06:2010, method A1S and fastness to rubbing was measured according to ISO 105-X12:2001.

- UV-protective properties: UVA transmittance (%), UVB transmittance (%) and Ultraviolet Protection Factor (UPF) of face side of dyed samples were measured according to the Australian/New Zealand standard AS/NZ4399/1996, using SDL Atlas Camspec M350 Ultraviolet-Visible Spectrophotometer. Based on the classification system suggested by the related standard, the fabrics were categorized according to their UPF ratings [43].

## 3. Result and Discussion

### 3.1 Results of Undyed Fabrics

In the current study three - undyed - fabrics were produced with three different weft settings (Table 1). As a consequence of having different settings, these fabrics exhibited different mass per unit area values (Table 2). However, thickness of these fabrics remained within a similar range.

The differences in structural properties may lead to significant changes in low-stress mechanical properties as well. It was observed that the fabric with the lowest weft setting (LL) has the lowest bending rigidity and the highest extensibility results (Table 3).

Higher deformation (D) and elasticity (E) results are usually desired for a better comfort. As can be seen in Table 4, sample ML has the highest D and E values. Similarly, ML has the lowest magnitude of permanent deformation (P) and hysteresis (H) when compared to samples LL and HL. These findings indicate that fabric ML can provide the optimum tactile comfort in terms of deformation and recovery behavior under out-of-plane deformation.

It was also observed that variations of fabric surface increased with the increase in weft setting (Table 5). For micro-surface variations (TS7) this finding was more prominent for face side results. On the other hand, the increase observed in macro-surface variations was more significant for back side of the fabrics. As higher surface variations are associated with a less desirable fabric surface, samples with lower TS7 and TS750 values may provide a more desirable tactile sensation.

The maximum pulling force recorded during nozzle test is commonly used as a measure of tactile comfort where a lower resistance when passing through a nozzle is an indicator of a better tactile comfort [41,42]. It was observed that sample LL has the lowest pulling force. Therefore, in terms of low-stress mechanical properties and nozzle test results, it is possible to state that LL exhibited the most desirable results.

A change in structural parameters may cause significant differences in permeability properties as well. It was recorded that LL has higher air permeability (Table 6). These findings indicated that using less number of weft yarns provided better tactile comfort, more desirable low-stress mechanical properties and better air permeability.

However, as warp yarns were made of 100% cotton and

weft yarns were made of 100% lyocell, using different weft settings provided differences in overall cotton/lyocell ratios of fabrics as well, which led to different water vapor permeability results. The transfer of water vapor from back side to face side of fabrics was faster for the fabric which has a moderate level of porosity and lyocell content (sample ML) (Table 6).

Moreover, fabrics with a high coverage may block more of the UV rays. In this respect, producing a fabric with low coverage and high porosity may lead to a less desirable level of UV protection. Results of the study indicated that fabrics with moderate and high weft settings (ML and HL) have similar levels of UV protection, which were higher than the sample with the lowest weft setting (LL) (Table 7). With respect to aforementioned test results, it was concluded that sample ML provided the optimum results. Therefore, sample ML was selected for the dyeing experiments.

Table 2. Structural properties of undyed fabrics

| Sample code | Mass per unit area | Thickness |
|---|---|---|
| | $(g/m^2)$ | $(\mu m)$ |
| LL | 123 | 254 |
| ML | 129 | 258 |
| HL | 135 | 256 |

Table 3. Low-stress mechanical properties and nozzle test results of undyed fabrics

| Sample code | Extensibility | Bending rigidity | Pulling force |
|---|---|---|---|
| | (%) | $(\mu J/m)$ | (cN) |
| LL | 2.40 | 7.38 | 25.2 |
| ML | 2.06 | 7.97 | 31.3 |
| HL | 1.99 | 8.37 | 37.4 |

Table 4. Out-of-plane deformation and recovery results of undyed fabrics

| Sample code | Deformation (D) | Elasticity (E) | Plasticity (P) | Hysteresis (H) |
|---|---|---|---|---|
| | (mm/N) | (mm/N) | $(\mu m)$ | (J) |
| LL | 1.81 | 1.63 | 175 | 89.9 |
| ML | 1.85 | 1.75 | 124 | 71.8 |
| HL | 1.81 | 1.66 | 146 | 77.9 |

Table 5. Surface characteristics of undyed fabrics

| Sample code | Side | Micro-surface variations (TS7) | Macro-surface variations (TS750) |
|---|---|---|---|
| LL | Face | 10.9 | 79.6 |
| ML | Face | 11.2 | 82.7 |
| HL | Face | 12.8 | 85.7 |
| LL | Back | 12.7 | 78.6 |
| ML | Back | 12.4 | 84.9 |
| HL | Back | 12.9 | 95.6 |

Table 6. Air permeability and water vapor permeability of undyed fabrics

| Sample code | Air permeability | Water vapor permeability |
|---|---|---|
| | (mm/s) | $(g/m^2/day)$ |
| LL | 100.6 ± 2.8 | 786.5 ± 20.2 |
| ML | 83.0 ± 2.4 | 813.2 ± 40.9 |
| HL | 60.7 ± 2.0 | 809.6 ± 33.2 |

Table 7. UV-protective properties of undyed fabrics

| Sample code | UVA | UVB | UPF | UPF rated |
|---|---|---|---|---|
| LL | 25.9 | 15.4 | 5.63 | 5 |
| ML | 25.4 | 14.7 | 5.86 | 5 |
| HL | 25.5 | 14.3 | 5.98 | 5 |

UVA: % UVA transmittance, UVB: % UVB transmittance, UPF: UV protection factor, UPF rated: Rated UV protection factor

### 3.2 Color Properties of Dyed Fabrics

It was observed that the color of the sample dyed without any mordant (ML1) was a moderate shade of yellowish pink (Figure 1). Sample ML2 - which was pre-mordanted with magnesium sulphate and simultaneously mordanted with ferrous sulphate - exhibited a brown color. Meanwhile pre-mordanting the fabric with ferrous sulphate resulted in a grey color (ML3). Results of color measurements indicated that color strength (K/S) was increased from 0.75 to 1.76 and 1.78 by the suggested mordanting processes (Table 8). On the other hand, mordanting caused a noticeable decrease in redness (a*), yellowness (b*) and chroma (C*) values.

All samples exhibited excellent dry rubbing fastness grades (5) and very good wet rubbing fastness grades (4-5) regardless of the mordanting option (Table 9). Washing fastness against staining was also at a very good level (≥4-5). The color change of unmordanted sample during washing fastness test was on a moderate level (3 and 2-3 for face and back sides, respectively), which was significantly improved after mordanting.

### 3.3 UV-Protective Properties of Dyed Fabrics

According to The Skin Cancer Foundation's Seal of Recommendation, a fabric must have a minimum Ultraviolet Protection Factor (UPF) of 30 to qualify as sun-protective clothing [44].
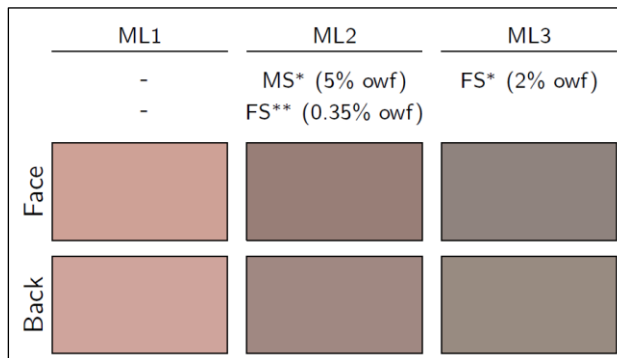


Figure 1. Color of face and back sides of the fabrics pre-mordanted* and simultaneously mordanted**with magnesium sulphate (MS) and ferrous sulphate (FS) and dyed with avocado seed extract

It was observed that sample ML1, which was dyed with avocado seed extract without any mordant has a UPF of 36.9 and a rated UPF of 35 (Figure 2.a), which indicates a very good level of protection against UV light. Even UVA and UVB transmittance values of ML1 slightly increased after washing, it was observed that ML1 still has a moderate level of UV protection after washing fastness test, as the rated UPF of ML1 after fastness test was 20 (Figure 2.b).

Both mordanted samples (ML2 and ML3) exhibited excellent levels of protection against UV light (UPF 50+). It was observed that sample ML3 which was pre-mordanted with ferrous sulphate exhibited the highest UPF result (Figure 2.a). The effective UV radiation transmission (%) values of ML2 and ML3 were lower than 2.5, and UPF values were 119 and 256, respectively.

The decrease in rated UPF of sample ML3 observed after washing fastness test was only 1.6% (Figure 2.b). For sample ML2, the decrease in UPF value after washing fastness test was 29.6%. Even after washing fastness test, the rated UPF of samples ML2 and ML3 remained as excellent (50+).

Table 8. Color results of samples dyed with avocado seed extract

| Sample Code | Side* | Mordant | K/S | L* | a* | b* | C* | h° |
|---|---|---|---|---|---|---|---|---|
| ML1 | Face | - | 0.751 | 70.2 | 16.0 | 12.2 | 20.1 | 37.3 |
| ML2 | Face | Mg, Fe | 1.763 | 55.0 | 9.61 | 7.50 | 12.2 | 38.0 |
| ML3 | Face | Fe | 1.781 | 55.8 | 4.03 | 4.84 | 6.30 | 50.2 |
| ML1 | Back | - | 0.653 | 71.4 | 15.1 | 10.7 | 18.5 | 35.3 |
| ML2 | Back | Mg, Fe | 1.333 | 58.7 | 8.84 | 6.46 | 11.0 | 36.2 |
| ML3 | Back | Fe | 1.789 | 58.8 | 3.84 | 7.18 | 8.14 | 61.9 |

*Side: Face side is warp dominant (cotton warp yarns are visible) and back side is weft dominant (lyocell weft yarns are visible).

Table 9. Washing, rubbing and light fastness results of fabrics dyed with avocado seed extract

| Fabric code | Side* | Washing fastness | | | | | | | Rubbing fastness | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Color change | Staining | | | | | | Dry | Wet |
| | | | WO | PAN | PET | PA 6.6 | CO | CA | | |
| ML1 | Face | 3 | 5 | 5 | 5 | 5 | 4-5 | 5 | 5 | 4-5 |
| ML2 | Face | 4-5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4-5 |
| ML3 | Face | 4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4-5 |
| ML1 | Back | 2-3 | 5 | 5 | 5 | 5 | 4-5 | 5 | 5 | 4-5 |
| ML2 | Back | 3-4 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4-5 |
| ML3 | Back | 4-5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 4-5 |

*Side: Face side is warp dominant (cotton warp yarns are visible) and back side is weft dominant (lyocell weft yarns are visible).

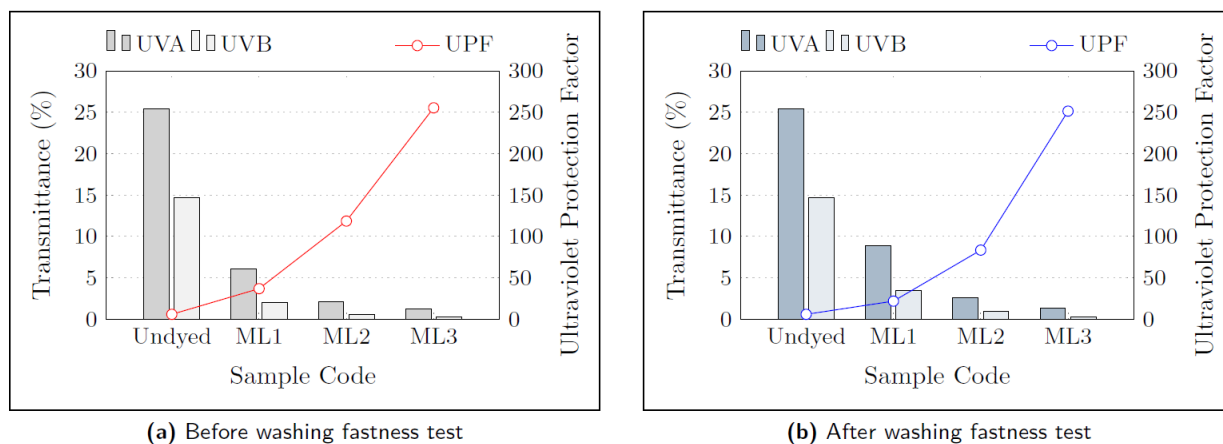**(a)** Before washing fastness test      **(b)** After washing fastness test

Figure 2. UVA (315 to 400 nm) and UVB (290 to 315 nm) transmittance (%) values and Ultraviolet Protection Factor (UPF) of fabrics dyed with avocado seed extract, measured (a) before and (b) after washing fastness test

## 4. Conclusions

Sunlight contains ultraviolet (UV) radiation which can negatively affect human skin in several ways. In this respect, developing alternative solutions which would provide protection against the possible harmful effects of UV rays is an important research topic.

In this study eco-friendly production steps for obtaining a UV-protective summer clothing with high clothing comfort were proposed. As lyocell is a more sustainable raw material when compared to cotton and a luxury fiber with several advantages such as high comfort and silk like structure, yarns made of 100% lyocell fibers were used as weft yarns and the effect of using different number of weft yarns on comfort and UV protection was discussed. Avocado seed was proposed as a natural dye source for textiles. Fabrics were mordanted with non-toxic metallic salts and dyed with avocado seed extract. Color, fastness, and UV-protective properties of dyed samples were determined and discussed.

Results of the study indicated that selecting different weft settings can provide significant changes in several fabric characteristics including low-stress mechanical properties and permeability. It was observed that both cotton and lyocell fibers had a good dyeability by the avocado seed extract, as both warp dominant face side and weft dominant back side of the dyed fabric samples exhibited vibrant colors with high K/S values and moderate to excellent fastness grades.

It was recorded that dyeing the studied fabric with avocado seed extract can significantly improve UV-protective properties as the UPF of raw fabric increased from 5.86 to 36.9 after dyeing, even without using any mordants. In the case of using different concentrations of magnesium sulphate and ferrous sulphate for mordanting, UPF was increased to an excellent level (50+) and these samples preserved their UV-protective property even after washing fastness test.

Findings of the study certified that it is possible to produce a light-weight fabric for summer clothing with excellent UV-protective properties and high clothing comfort, using sustainable processes, without including any toxic materials.

## Declaration

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

## Author Contributions

N. Uren developed the methodology, performed the experiments, measurements and data analysis, and wrote the manuscript.

## References

1. American Cancer Society, [cited 2023 01 October]; Available from: https://www.cancer.org/cancer/risk-prevention/sun-and-uv/uv-radiation.html

2. Alfredsson, L, B.K. Armstrong, D.A. Butterfield, R. Chowdhury,F.R. de Gruijl, M. Feelisch, C.F. Garland, P.H. Hart, D.G. Hoel, R. Jacobsen et al., *Insufficient Sun Exposure Has Become a Real Public Health Problem*. International Journal of Environmental Research and Public Health, 2020. **17**(14): 5014.

3. Lucas, R.M., and A.L. Ponsonby, *Considering the potential benefits as well as adverse effects of sun exposure: Can all the potential benefits be provided by oral vitamin D supplementation?*. Progress in Biophysics and Molecular Biology, 2006. **92**(1): p. 140-149.

4. Rhee, H.J. van der, E. de Vries, and J.W. Coebergh, *Regular sun exposure benefits health*. Medical Hypotheses, 2016. **97**: p. 34-37.

5. Scott, R.A. *Cold weather clothing for military applications*. In: Textiles for Cold Weather Apparel, Edited by Williams, J.T. Cambridge UK: Woodhead Publishing, 2009, p. 305-328.

6. Gokarneshan, N. *Comfort properties of textiles: A review of some breakthroughs in recent research*. Juniper Online Journal of Material Science, 2019. **5**(3): 555662.

7. Nazir, M.U., K. Shaker, Y. Nawab, S.T.A. Hamdani, H.M. Abdullah, and M. Umair, *Thermo-physiological comfort of woven fabrics made from different cellulosic yarns*. Journal

of Natural Fibers, 2022. **19**(11): p. 4050-4062.

8. Rasheed, A., M.H. Malik, F. Ahmad, F. Azam, and S. Ahmad, *Effect of fibers and weave designs on the thermo-physiological comfort of summer scarf fabric*. Advances in Materials Science and Engineering, 2022. **2022**: 8384193.

9. Lee, K.P., J. Yip, and K.L. Yick, *Investigating the factors affecting the thermal and tactile comfort of summer undergarments*. Human Factors for Apparel and Textile Engineering, 2022. **32**: p. 46-55.

10. Kibria, G., M.R. Repon, M.F. Hossain, T. Islam, M.A. Jalil, M.D. Aljabri, and M.M. Rahman, *UV-blocking cotton fabric design for comfortable summer wears: factors, durability and nanomaterials*. Cellulose, 2022. **29**: p. 7555-7585.

11. Shah, M.A., B.M. Pirzada, G. Price, A.L. Shibiru, and A. Qurashi, *Applications of nanotechnology in smart textile industry: A critical review*. Journal of Advanced Research, 2022. **38**: p. 55-75.

12. Majumdar, A., V.K. Kothari, and A.K. Mondal, *Engineering of cotton fabrics for maximizing in vitro ultraviolet radiation protection*. Photodermatology, Photoimmunology & Photomedicine, 2010. **26**(6): p. 290-296.

13. Feng, X.X., L.L. Zhang, J.Y. Chen, and J.C. Zhang, *New insights into solar UV-protective properties of natural dye*. Journal of Cleaner Production, 2007. **15**(4): p. 366-372.

14. Grifoni, D., L. Bacci, G. Zipoli, L. Albanese, and F. Sabatini, *The role of natural dyes in the UV protection of fabrics made of vegetable fibres*. Dyes and Pigments, 2011. **91**(3): p. 279-285.

15. Karabulut, K., and R. Atav, *Dyeing of cotton fabric with natural dyes without mordant usage Part I: Determining the most suitable dye plants for dyeing and UV protective functionalization*. Fibers and Polymers, 2020. **21**: p. 1773-1782.

16. Otaviano, B.T.H., M. Sannomiya, F.S. de Lima, M.M.P. Tangerina, C.I. Tamayose, M.J.P. Ferreira, and S.M. da Costa, *Pomegranate peel extract and zinc oxide as a source of natural dye and functional material for textile fibers aiming for photoprotective properties*. Materials Chemistry and Physics, 2023. **293**: 126766.

17. Hou, X., X. Chen, Y. Cheng, H. Xu, L. Chen, and Y. Yang, *Dyeing and UV-protection properties of water extracts from orange peel*. Journal of Cleaner Production, 2013. **52**: p. 410-419.

18. Benli, H., and M.I. Bahtiyari, *Testing acorn and oak leaves for the UV protection of wool fabrics by dyeing*. Journal of Natural Fibers, 2022. **19**(14): p. 7925-7938.

19. Alebeid, O.K., Z. Tao, and A.I. Seedahmed, *New approach for dyeing and UV protection properties of cotton fabric using natural dye extracted from henna leaves*. Fibres & Textiles in Eastern Europe, 2015. **5**(113): p. 60-65.

20. Uren, N., *Use of kermes oak (Quercus coccifera L.) seed as a sustainable source for uv-protective functionalization of textiles*. 5th International Black Sea Modern Scientific Research Congress, Rize, Turkey, 2023. **1**: p. 685-692.

21. Uren, N., *Eco-friendly coloration and UV-protective functionalization of textiles using mango (Mangifera indica L.) peel and seed extracts*. 6th International Antalya Scientific Research and Innovative Studies Congress, Antalya, Turkey, 2023. p. 278-285.

22. Ellams, D.L., Christie, R.M., and Robertson, S., *An approach to sustainable coloration of lyocell fabrics by screen printing using extracts of leaves and bark from eucalyptus*. Coloration Technology, 2014. **130**(1), p. 48-53.

23. Tayyab, N., Javeed, A.A., Sayed, R.Y., Mudassar, A., Faisal, R., Ahmad, F., and Muhammad, A., *Dyeing and colour fastness of natural dye from Citrus aurantium on Lyocell fabric*. Industria Textila, 2020. **71**(4), p. 350-356.

24. Banu, M.N., and Fathima, M.S., *A comparative study on the color fastness properties of Lyocell and Bamboo fabrics after dyeing with a natural dye: Chavalkodi*. International Journal of Home Science, 2018. **4**(1), p. 37-44.

25. Ke, G., Zhang, J., Li, W., and Liu, K., *Over dyeing of Lyocell fabric with natural indigo and rhubarb*. Textile Research Journal, 2023. **93**(23-24), p. 5437-5447.

26. Ke, G., Mulla, M.S., Peng, F., and Chen, S., *Dyeing properties of natural Gardenia on the lyocell fabric pretreated with tannic acid*. Cellulose, 2023. **30**(1), p. 611-624.

27. Birhanlı, A., and M. Özmen, *Evaluation of the toxicity and teratogenity of six commercial textile dyes using the frog embryo teratogenesis assay-Xenopus*. Drug and Chemical Toxicology, 2005. **28**(1), p. 51-65.

28. Eroğlu Doğan, E., E. Yeşilada, L. Özata, and S. Yoloğlu, *Genotoxicity testing of four textile dyes in two crosses of drosophila using wing somatic mutation and recombination test*. Drug and Chemical Toxicology, 2005. **28**(3): p. 289-301.

29. Basit, A., W. Latif, S.A. Baig, and A. Afzal, *The mechanical and comfort properties of sustainable blended fabrics of bamboo with cotton and regenerated fibers*. Clothing and Textiles Research Journal, 2018. **36**(4): p. 267-280.

30. Jiang, X., Y. Bai, X. Chen, and W. Liu, *A review on raw materials, commercial production and properties of lyocell fiber*. Journal of Bioresources and Bioproducts, 2020. **5**(1): p.16-25.

31. Adnan, M., and J. Moses, *A study on the thermophysiological and tactile comfort properties of silk/lyocell blended fabrics*. Revista Materia, 2020. **25**(03): p.

32. Edgar, K. J., and H. Zhang, *Antibacterial modification of lyocell fiber: A review*. Carbohydrate Polymers, 2020. **250**: 116932.

33. Food and Agriculture Organization of the United Nations, [cited 2024 21 February]; Available from: https://www.fao.org/faostat/en/#data/QCL

34. Uren, N. *Eco-friendly dyeing of cotton and wool fabrics with avocado seed and peel extracts*. Journal of Natural Fibers, 2022. **19**(16): p. 13765-13775.

35. Uren, N., and Kutlu, B., *Natural dyeing of plasma treated wool with avocado seed extract and use of tartaric acid as bio-mordant*. Coloration Technology, 2024. Early access.

36. Cuk, N. and M. Gorjanc, *Natural dyeing and UV protection of raw and bleached/mercerised cotton*. Tekstilec, 2017. **60**(2): p. 126-136

37. Tyagi, G.K., Bhattacharyya, S., Bhowmick, M., and Narang, R., *Study of cotton ring-and compact-spun yarn fabrics: Part II–Effects of spinning variables on comfort characteristics*. Indian Journal of Fibre & Textile Research, 2010. **35**(2), p. 128-133.

38. Uren, N., *Determining deformation and recovery characteristics of woven fabrics using a novel instrument*. Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering, 2024. **26**(76): p. 90-97.

39. Turkoglu, G.C., and Uren, N., Investigating comfort components of non-woven surfaces suitable for the skin

layer of sanitary pads. Kahramanmaras Sutcu Imam University Journal of Engineering Sciences, 2023. **26**(4): p. 922-931

40. Pan, N., and K.C. Yen, *Physical interpretations of curves obtained through the fabric extraction process for handle measurement*. Textile Research Journal, 1992. **62**(5): p. 279-290.

41. Uren, N., and A. Okur, *Analysis and improvement of tactile comfort and low-stress mechanical properties of denim fabrics*. Textile Research Journal, 2019. **89**(23-24): p. 4842-4857.

42. Sular, V., and A. Okur, *Objective evaluation of fabric handle by simple measurement methods*. Textile Research Journal, 2008. **78**(10): p. 856-868.

43. Algaba, I., and A. Riva, *In vitro measurement of the ultraviolet protection factor of apparel textiles*. Coloration Technology, 2002. **118**(2): pp. 52-58.

44. Skin Cancer Foundation, [cited 2023 01 October]; Available from:

https://www.skincancer.org/skin-cancer-prevention/seal-of-recommendation/.

*Research Article*

# Constructing a descriptive sensory panel for tactile comfort evaluations: Effect of demographic variables and panel size

*Nazli Uren* [a,*] (iD)

*aDepartment of Textile Engineering, Dokuz Eylul University, Izmir 35390, Turkey*

| ARTICLE INFO | ABSTRACT |
|---|---|
| | Sensory tests are essential components of comfort studies, and constructing a sensory panel is a crucial step of this process. In the current study, Total Hand (TH) scores of 41 woven fabrics were determined by assessors having different demographic characteristics. Assessment accuracy and inter-rater reliability of panel members were investigated via correlation and concordance analyses. Effect of demographic variables (gender, age, and level of expertise), panel size and sampling method on sensory evaluation results were discussed based on statistical measures. Findings of the study certified that sensory evaluations carried out with female panel members represent overall TH scores more successfully than males and assessment of female participants are in a better agreement with each other. It was also observed that assessment accuracy and inter-rater agreement improved with increasing levels of expertise. Investigations revealed that small panel sizes were sufficient to accurately evaluate fabric hand. Therefore, it was concluded that increasing the number of participants may not necessarily provide further information on comfort preferences and perceptions of potential customers. |

## 1. Introduction

Comfort is an important aspect which determines the commercial value of textile products. Many studies have been conducted in recent decades to establish a reliable test method which will accurately determine tactile comfort of textiles [1-2]. There are several measurement systems and devices designed for objective measurement of fabric hand [3-9]. In many cases, it is essential to apply both objective and subjective approaches to obtain verified results. In this respect, sensory tests are considered as the fundamentals of comfort studies. Even though a number of standards and procedures were proposed to introduce the methodology of different sensory analysis techniques such as paired comparisons, duo-trio tests and triangle tests, researchers might also prefer to follow a custom evaluation method which will correspond better with the time, labor, material and financial constraints [10-14]. In both cases, decisions regarding sampling method, panel size and demographics of panel members should be made on the grounds of scientific knowledge [15].

There are different sampling methods which can be used when constructing a sensory panel such as convenience sampling, random sampling, and purposive sampling. Convenience sampling is a non-probability sampling method where participants are selected for inclusion in the sensory panel because they are the easiest for the researcher to access. For instance, researchers who investigate tactile comfort of fabrics via sensory evaluations usually prefer to work with university students or work associates due to reasons such as geographical proximity, availability at a given time, or willingness to participate in the research. Even though commonly preferred by researchers, in non-probability sampling methods, the panel may not represent each member of the population. Therefore, these methods are considered to be less objective than probability (random) sampling techniques [16].

Tactile sensitivity of humans may vary depending on several variables such as physical condition, age, gender etc. The somatosensory system is a network of neural structures, and the conscious perception of touch is a cooperation of several body parts including the skin, spinal cord, and the brain. Consequently, in addition to tactile sensitivity, the transformation of sensory signals into a conscious perception may vary depending on factors

including knowledge of the research issue, product usage behavior, physical and personal characteristics, demographics, psychographics and observed behaviors [17,18].

Demographic segmentation includes characteristics such as gender, age, education, income, ethnicity and physical attributes. The possible differences among evaluations of assessors from different countries and nationalities were previously investigated by several researchers [19-22]. A good agreement among sensory evaluations of assessors from Japan, Australia, New Zealand, and India was recorded in the case of fabric hand being ranked on a simple scale according to assessors' own interpretation of fabric hand [19]. In another study, it was concluded that assessors from different cultures may interpret some terms differently if complicated expressions were used [20].

Besides the effect of ethnicity and culture, conditions such as temperature and humidity which are related to geographic location may also affect the tactile sensation. For instance, Speijers et al. [21] detected significant differences between discomfort sensations of Australian and Chinese wearers for trials carried out in hot environment during high levels of activity when wearing wool and cashmere products. However, Keefe et al. [22] stated that the difference between ethnic and cultural groups' pain behavior is often smaller than the effect of age and gender.

Marketing research advisors suggest that it is beneficial to include the opposite sex in the consumers' studies, even if the evaluated material is mostly used by one gender only [17]. A review of the previous literature indicated that researchers commonly prefer to work with female assessors during sensory evaluations [23-28]. In fact, in some studies, all panel members were females [21,29-32]. Effect of gender differences has been investigated in a variety of sensory tests [33-37], which mostly revealed that females have a greater tactile detection sensitivity than males. Roh et al. [38] stated that gender, age, and professionalism affect the perception of sensation and suggested the more females than males, the lower the age of assessors, and the more experts than nonexperts, the more sensitive their feeling would be in a subjective evaluation.

Researchers might prefer to work with untrained consumers and members of a local community as panel members, or they may prefer to work with expert assessors - university students, research assistants, lecturers, researchers, and workers of a company - to acquire a comprehensive and prudent data. In a study realized by Asad et al. [39], prickle sensation of fabrics was evaluated by trained university students. The inter-rater reliability reported for these assessors was at a moderate level. In another study realized by Harpa et al. [40], a minimum

training was given to second year students of textile bachelor level who were regarded as untrained consumers. The level of agreement reported for these assessors was also quite low.

Researchers suggested categorization of assessors based on their electroneurophysiological responses to fabric-skin contact [41]. Hui et al. [26] worked with assessors selected based on their interest, availability, and successful completion of a tactile sensitivity screening test, and they reported a good test-retest reliability for the selected assessors. Similarly, Musa et al. [42] proposed a method to select assessors based on their level of finger sensitivity to minimize the disagreements among the panel members due to demographic criteria (age, gender, origin). Xue et al. [43] emphasized that simple sensory experiments such as sorting and rating tasks can be consistently performed by assessors without any specific expertise or training.

When constructing a panel for sensory evaluations, it is important to define a reasonable age range as well. If a research study does not specify a relevant gender, age, education, income, or ethnicity, it is suggested to construct a panel with assessors that covers as many demographic characteristics as possible [17]. The study of United Nations regarding world population prospects indicated that the global median age has increased from 21.5 to over 30 years, and the global population breakdown by age showed that half of the world population is between 25 and 65 [44]. The population reports also pointed that higher-income countries across North America, Europe and East Asia have a higher age median. In this respect, including older members in a sensory panel may be useful to gain a better understanding of the consumers' preferences. However, researchers who investigated fabric hand commonly preferred to work with university students or young adults [21,25-31,39,40,43,45-47], and participants with ages 40 and above were rarely included in sensory evaluations [24,42,48,49].

Considering the time and labor factors, it is important to determine the adequate panel size which would provide accurate, reliable, and reproducible sensory evaluation results. The number of panel members should be sufficient to produce reliable and representative sensory data. On the other hand, the involvement of a large number of assessors may cause an increase in time consumptions and cost [43]. Therefore, panel size should be limited to maintain efficient working conditions. Researchers who carried out sensory evaluations by trained or expert assessors may prefer to work with small panel sizes [23,24,48,50,51]. Xue et al. [43] stated that nonexpert assessors - when compared to trained or expert assessors - can be less capable of making a bias-free assessment in pure sorting tasks due to their limited knowledge on sensory evaluations. Therefore, it was suggested to work with a larger number of assessors with better representativeness

for studies with the objective of acquiring comprehensive and prudent data. Stanton et al. [32] carried out wearer trials with 43 assessors and concluded that a reduction to 25 wearers was adequate for later trials with minimal loss in sensitivity.

A review of the previous literature indicated that researchers commonly preferred convenience sampling method for sensory evaluation of fabric hand. During construction of a descriptive sensory panel, if the members are selected from a closer circle but not from a large and more diverse population, decisions regarding demographic characteristics and panel size emerge as important factors. In the current study, the effects of gender, age, level of expertise and panel size on accuracy of sensory evaluations - in case of constructing the sensory panel according to convenience sampling method - were investigated. The effect of selecting a non-probability sampling method on the statistical findings was also discussed. As a secondary purpose, a sensory evaluation procedure which would be applicable regardless of assessors' ethnicity or cultural differences was proposed.

## 2. Materials and Methods

### 2.1 Material

The study was carried out with 41 conventional woven fabrics made of cotton and its blends with various fibers such as polyester, viscose, angora, wool, silk, lyocell, elastane and linen. The studied fabrics have plain, 2/1 twill or 3/1 twill weave patterns and a mass per unit area between 150 and 275 $g/m^2$.

### 2.2 Sensory Panel

Sensory tests were realized with participation of 200 assessors having different demographic characteristics. Panel members were chosen according to convenience sampling method. The panel members were students and academic staff of faculty of engineering (textile, mechanical, electrical-electronic, industrial and materials etc.), education, human sciences, social sciences, management, maritime and fine arts of the university at which the research took place.

In addition to these participants, students and academic staff of another local university, and workers (having different levels of education) of a textile manufacturer were also included in the panel. Any person over the age of 18 and willing to participate in the sensory evaluations was included in the study. No additional selection process was carried out.

The effect of age on sensory evaluation results was investigated for three sub-groups: youngsters, young adults, and adults (Table 1). Panel members were classified into three age clusters by a two-step cluster analysis. Segmentation for age clusters was based on age

variable only, and the number of clusters was fixed to three.

The difference between total hand evaluations of assessors with different levels of expertise was investigated for four sub-groups. Expert assessors were researchers who have a bachelor's or higher degree in textile engineering. Qualified assessors were workers of textile related majors - technicians, laboratorians, or designers - with degrees other than textile engineering. Novice assessors were undergraduate students of the textile engineering faculty. Nonexpert assessors were workers and students from any major except textiles.

### 2.3 Sensory Evaluation Procedure

Fabric hand assessments were made based on the total hand feel of fabrics. To increase the accuracy of sensory evaluations and prevent context, order and position effects, two specimens were prepared in warp and weft directions for each fabric type, and a mixed batch of specimens was presented to the assessor with a random positioning and order. To exclude bias caused by fabric appearance, assessments were carried out according to blind test requirements.

Correct lexicon usage is the ability of an assessor to understand and use attributes in a similar manner [18]. Differences in lexicon usage may lead to agreement issues when if one assessor understands the attribute to mean something different from the other panel members [19,22]. In this respect, it is important that each attribute being assessed has a precise definition which can be clearly understood by all assessors. To eliminate the bias caused by lexicon usage, a five-point hedonic scale with simple and universally understandable labels was used in the current study. The assessors were asked to arrange fabric samples into five groups, in the order of a five-point scale. The scale points were introduced as; 5 excellent, 4 good, 3 average, 2 fair, and 1 poor hand quality. At the end of the evaluation session, the numbers associated with the indicated groups were regarded as Total Hand (TH) scores and the score of each sample was recorded.

Assessors evaluated the fabric hand freely, according to their preferences and the attributes they prioritize the most. Sensory tests were performed as single trials. In order to not interfere with the original decisions of assessors, no training or preparation trial was carried out.

### 2.4 Assessment Accuracy and Inter-Rater Agreement

Assessment accuracy of panel members was determined based on the assumption that the overall TH scores determined by participation of 200 assessors represent the opinion of the population on total hand of investigated fabrics. The closeness between the total hand values estimated by an individual participant and the overall TH scores (obtained from 200 participants) was statistically measured by Spearman's rank order correlation analysis and regarded as a measure of assessment accuracy.

Table 1. Demographic segmentation of sensory panel

| Demographics | Sub-group | Description | Panel size | | |
|---|---|---|---|---|---|
| | | | All | Female | Male |
| Gender | Female | Female | 100 | 100 | - |
| | Male | Male | 100 | - | 100 |
| Age | Youngster | Age between 19-26 | 114 | 64 | 50 |
| | Young adult | Age between 27-42 | 75 | 31 | 44 |
| | Adult | Age between 43-66 | 11 | 5 | 6 |
| Expertise | Nonexpert | Workers and students of non-textile related majors | 62 | 25 | 37 |
| | Novice | Undergraduate textile engineering students | 76 | 48 | 28 |
| | Qualified | Technicians, laboratorians, designers | 22 | 7 | 15 |
| | Expert | Textile engineers with bachelor's degree or higher | 40 | 20 | 20 |

Assessment accuracy of sub-groups was estimated using correlation coefficients calculated for panel members in that sub-group. The significance of the differences observed between accuracy of different sub-groups was evaluated at 90% and 95% confidence levels, using Mann-Whitney and Kruskal-Wallis Tests. To summarize the findings regarding assessment accuracy, panel members were classified into three clusters based on the correlation coefficient data, using statistical software.

Kendall's coefficient of concordance is often used to determine the agreement among members of a sensory panel, where a higher Kendall's W value indicates a better inter-rater agreement and it is a measure of how much homogeneity or consensus there is in the scores given by panel members [39,40]. In the current study, inter-rater agreement was discussed based on Kendall's W values.

The repeatability of a panel member is referred as intra-rater reliability, and it is a common indicator of assessment accuracy. In the current study, all assessors evaluated the investigated fabrics in a single trial where no second trial was carried out. In this respect, repeatability measures were not included in the study.

## 3. Result and Discussion

### 3.1 Effect of Demographic Variables on Assessment Accuracy

Statistical findings certified that assessments of female participants were more correlated with overall TH scores than males (Table 2), and this difference was statistically significant (p=0.000). It was also observed that accuracy of panel members improves by expertise (p=0.003). Even though the correlation coefficients calculated for adult and young adult participants were considerably higher than youngsters, age was found out to be the least efficacious characteristic among the investigated demographic variables (p=0.091). When the effect of age and expertise on correlation coefficients was investigated separately for female and male participants, it was detected that correlation relations drastically improved by increasing the age and level of expertise of female assessors. On the other hand, for male assessors, only a slight improvement was observed in terms of expertise (Table 2).

To summarize the findings regarding assessment accuracy, panel members were classified into three clusters. As can be seen in Figure 1, 140 out of 200 panel members have very high accuracy, whereas assessments of 49 members were moderately accurate. When the number of assessors in accuracy clusters was investigated in terms of demographic variables, it was observed that 75% of female assessors have high accuracy, while this number was 65% for male assessors (Figure 2). Results of cluster analyses also indicated that expertise has a great effect on accuracy. In fact, 80% of expert assessors were in high accuracy cluster, while this number was between 66% and 68% for assessors with limited or no expertise.

Table 2. Assessment accuracy of panel members in different demographic sub-groups

| Demographics | Spearman's rho* | | |
|---|---|---|---|
| | ALL | Female | Male |
| Female | 0.793 ± 0.133 | 0.793 ± 0.133 | - |
| Male | 0.742 ± 0.140 | - | 0.742 ± 0.140 |
| Youngster | 0.755 ± 0.133 | 0.763 ± 0.149 | 0.744 ± 0.110 |
| Young adult | 0.785 ± 0.151 | 0.847 ± 0.079 | 0.740 ± 0.174 |
| Adult | 0.784 ± 0.093 | 0.841 ± 0.044 | 0.738 ± 0.100 |
| Nonexpert | 0.738 ± 0.160 | 0.752 ± 0.162 | 0.729 ± 0.160 |
| Novice | 0.762 ± 0.123 | 0.770 ± 0.131 | 0.748 ± 0.108 |
| Qualified | 0.762 ± 0.155 | 0.826 ± 0.069 | 0.732 ± 0.176 |
| Expert | 0.827 ± 0.108 | 0.888 ± 0.044 | 0.765 ± 0.118 |
| All | 0.767 ± 0.139 | 0.793 ± 0.133 | 0.742 ± 0.140 |

*Correlation coefficient was calculated between assessment of an individual panel member and overall Total Hand (TH) scores and presented as average ± standard deviation.
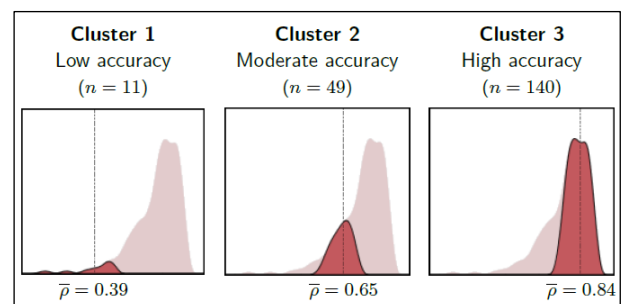


Figure 1. Assessment accuracy clusters of panel members constructed based on correlation coefficients (Spearman's rho, ρ) calculated between assessment of each individual panel member and overall Total Hand (TH) scores

Based on these findings it was concluded that most of the assessors have a moderate to good level of accuracy in general, regardless of demographics. However, sensory evaluations carried out with females, experts, young adults, and adults provided a better accuracy.

### 3.2 Effect of Demographic Variables on Scale Usage

Using similar ratings across all samples may indicate a low sensory acuity [18]. Poorly discriminating assessors may prefer to use a "safe scale range" to cover their inability to discern the evaluated attribute. In Figure 3, the distribution of TH scores determined by assessors in different demographic sub-groups were presented in the form of box-plots. These plots indicated that females, adults, and experts used a larger portion of the 5-point scale when compared to the assessors in other demographic sub-groups.
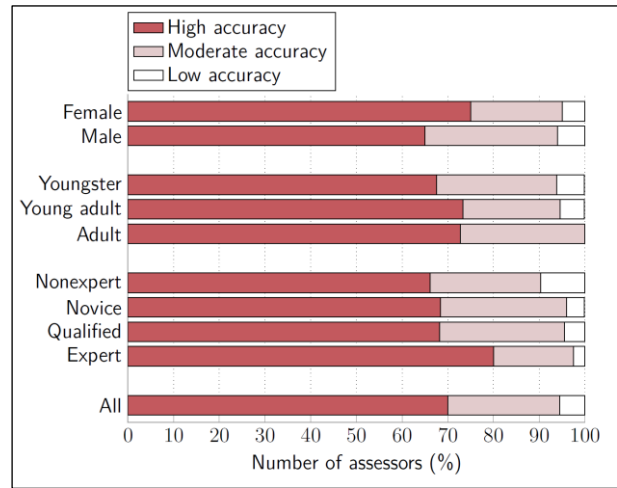


Figure 2. Number of assessors in accuracy clusters constructed based on correlation coefficients (Spearman's rho, $\rho$) calculated between assessment of each individual panel member and overall Total Hand (TH) scores
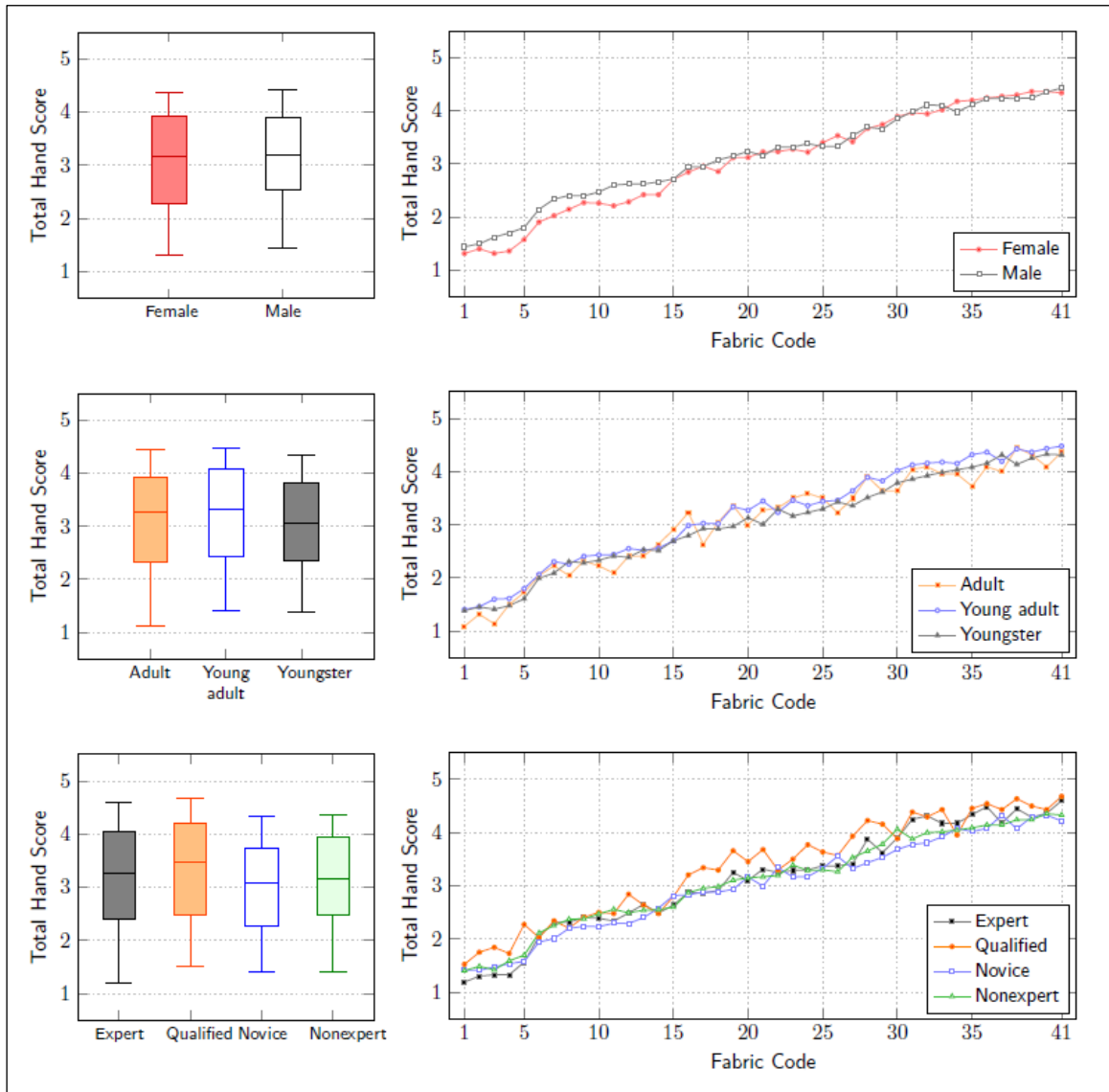


Figure 3. Total Hand scores of fabrics evaluated by assessors in different demographic sub-groups

TH score of each investigated fabric were given in Figure 3. TH scores were mostly within a similar range, regardless of demographics. In accordance with the differences recorded in scale usage behaviors, several differences in TH scores evaluated by participants in different demographic sub-groups were observed. For instance, fabrics with average or higher ($\geq$3) hand qualities were rated with similar TH scores by female and male assessors. On the other hand, fabrics with fair or poor hand qualities (<3) were rated with lower TH scores by females and higher TH scores by males (Figure 3). The differences calculated between TH scores determined by female and male assessors were less than 0.40 points on a five-point scale.

The variations in distribution of TH scores attained by assessors in different age sub-groups were smaller than the variations recorded for expertise sub-groups, and larger than the variations recorded for gender sub-groups. It was recorded that young adults rated fabrics with relatively higher TH scores.

Noticeable differences were observed among assessments of participants in different expertise sub-groups as well. Qualified assessors rated most of the fabrics with higher TH scores. On the contrary, novice assessors rated the fabrics with considerably lower TH scores. In fact, the highest difference calculated among TH score assessments of different expertise sub-groups were recorded between novice and qualified assessors, which was 0.79 points. It was also found out that fabrics with better tactile comfort were rated with higher TH scores and fabrics with less desirable levels of tactile comfort were rated with lower TH scores, when evaluated by expert assessors.

### 3.3 Effect of Demographic Variables on Inter-Rater Agreement

Kendall's coefficient of concordance is a measure of agreement among panel members. Calculated concordance coefficients certified that female assessors were in a better agreement with each other than males (Table 3). Kendall's W values indicated that the agreement among assessors improved by age and expertise, and this effect was stronger for females. These findings were also in accordance with the results of correlation analyses presented in section 3.1.

### 3.4 Effect of Panel Size and Sampling Method on Assessment Accuracy

For panel size investigations, panels with different sizes were constituted regardless of their demographic characteristics. Panel members of the current study were selected according to convenience sampling method where the readily approachable panel members are selected for participation. To investigate the effect of sample size - in

case of convenience sampling - the participants were grouped according to their order of participation, and the effect of panel size on accuracy was investigated in terms of correlation coefficients calculated between assessments of individual panel members and overall TH scores (Table 4). When the panels constructed with convenience sampling method were investigated, it was observed that the evaluation accuracy of the first 10 assessors of the study was the highest among all investigated panel sizes. As the number of participants increased from 10 to 100, the accuracy of the panel gradually decreased. Meanwhile the accuracy of panels with 100 or more assessors was almost the same ($\rho$=0.77).

Convenience sampling is a non-probability sampling method commonly preferred by researchers during participant selection for sensory evaluations. However, it is claimed that using a non-probability sampling method might have several possible effects on sensory evaluation results [16]. To investigate the effect of sampling method on accuracy of panels with different sizes, accuracy of panel members - in case of using probability (random) sampling method - was also investigated.

Table 3. Inter-rater agreement of demographic sub-groups

| Demographics | Kendall's W* | | |
|---|---|---|---|
| | ALL | Female | Male |
| Female | 0.638 | 0.638 | - |
| Male | 0.560 | - | 0.560 |
| Youngster | 0.578 | 0.592 | 0.571 |
| Young adult | 0.627 | 0.739 | 0.559 |
| Adult | 0.649 | 0.770 | 0.618 |
| Nonexpert | 0.557 | 0.587 | 0.544 |
| Novice | 0.592 | 0.608 | 0.584 |
| Qualified | 0.606 | 0.717 | 0.582 |
| Expert | 0.705 | 0.814 | 0.625 |
| All | 0.596 | 0.638 | 0.560 |

*Asymptotic significance is equal to 0.000 for all Kendall's W values.

Table 4. Assessment accuracy of panel members calculated for different panel sizes

| Panel size | Spearman's rho* | |
|---|---|---|
| | Convenience sampling | Random sampling |
| 5 | 0.818 ± 0.078 | 0.851 ± 0.102 |
| 10 | 0.845 ± 0.067 | 0.757 ± 0.208 |
| 15 | 0.829 ± 0.074 | 0.776 ± 0.148 |
| 20 | 0.834 ± 0.079 | 0.769 ± 0.125 |
| 25 | 0.823 ± 0.086 | 0.761 ± 0.155 |
| 50 | 0.803 ± 0.104 | 0.787 ± 0.106 |
| 75 | 0.785 ± 0.122 | 0.761 ± 0.138 |
| 100 | 0.767 ± 0.148 | 0.784 ± 0.123 |
| 125 | 0.766 ± 0.141 | 0.762 ± 0.145 |
| 150 | 0.766 ± 0.140 | 0.780 ± 0.129 |
| 175 | 0.775 ± 0.135 | 0.768 ± 0.142 |
| 200 | 0.767 ± 0.139 | 0.767 ± 0.139 |

*Correlation coefficient was calculated between assessment of an individual panel member and overall Total Hand (TH) scores, and presented as average ± standard deviation.

To investigate the effect of panel size in case of using probability sampling method, random sampling conditions were generated with the help of statistical software and the desired number of panel members was selected among 200 participants according to random sampling principle. When the correlation coefficients calculated between assessments of randomly selected panel members and overall TH scores were investigated, it was observed that the panel members have the highest accuracy when the panel size was equal to five (Table 4). This finding indicated that each member of a panel with five assessors selected randomly among 200 participant candidates has a higher probability to represent the overall assessment of 200 assessors. Moreover, when the panel size increases, the possibility to include assessors with relatively lower accuracies increases as well. The accuracy of panels constructed with convenience sampling principle indicated a similar trend. Yet the decrease in accuracy caused by the increase in panel size was less drastic in case of using convenience sampling method.

### 3.5 Effect of Panel Size and Sampling Method on Inter-Rater Agreement

Kendall's concordance coefficients were calculated for panels with different number of participants, and the level of agreement among panel members were investigated for panels with different sizes, constructed according to convenience sampling and random sampling methods. Larger panel sizes offer a wider diversity of assessors with possible differences regarding preferences on fabric hand. Accordingly, it is predictable to observe a lower level of agreement among assessors of larger panels. Concordance results of panels constructed with convenience sampling method indicated that the level of agreement among the first 10 assessors of the study was the highest (W=0.771). Similarly, the panel with five randomly selected assessors exhibited the highest Kendall's W value (W=0.758) (Figure 4).
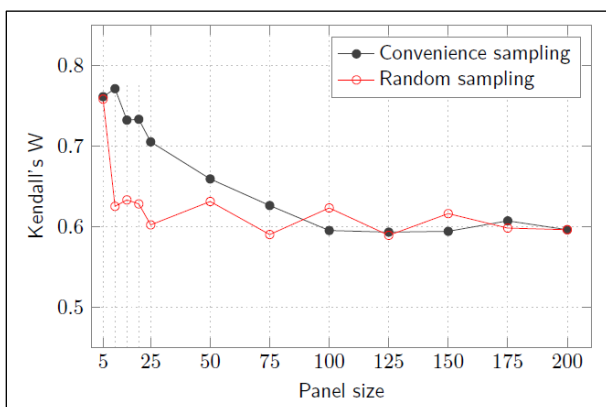


Figure 4. Kendall's coefficient of concordance (Kendall's W) values calculated for different panel sizes as a measure of inter-rater agreement

When the collection of new data does not provide any further information on the issue under investigation, it is usually concluded that the panel size has reached its saturation point [53,54]. For convenience sampling, when the panel size exceeded 100, the calculated Kendall's W values were almost constant (0.60 ± 0.01) (Figure 4). Based on these findings, it was concluded that a panel size of 100 is the saturation point for sensory evaluation of fabric hand. In the case of selecting assessors according to random sampling method, similar levels of agreement were recorded for panels with 10 or more assessors (0.61 ± 0.02), thus - in such case - the saturation point was reached with 10 assessors. This finding indicated that the panel size may reach the saturation point earlier, depending on the sampling method.

### 3.6 Discussion

Tactile properties of textiles have a great effect on purchase decision of consumers and fabric hand is one of the important parameters that determine the market value of a textile product. Researchers have been investigating this phenomenon for several decades and numerous objective measurement methods were proposed for determining hand related properties of textiles. However, the significance of sensory evaluations was not altered by the developments in measurement technologies.

Most of the current standards proposing guidelines for sensory evaluations are compatible with majors other than textiles. Therefore, researchers may prefer to carry out a custom sensory evaluation technique instead of following a given standard. In such cases, decisions regarding selection of panel members, scale type and test protocol must be carefully considered, as these parameters may impact the accuracy of sensory evaluation results.

In the current study, hand of conventional woven fabrics was evaluated by 200 panel members and the assessment performance of participants with different demographic characteristics was investigated. The participants of the study were coming from the same geographic location, belonging to a similar culture and members of the same nationality. Therefore, effect of these parameters on sensory evaluation results was not discussed in this report. It was aimed to achieve scientific findings which may represent a larger and more diverse population. Therefore, in the current study, a basic five-point scale with simple and universally relatable labels was used.

It was observed that 94.5% of panel members have a moderate or good level of accuracy. Female participants exhibited a more desirable level of accuracy, and this accuracy was further improved by increasing age and expertise. Even though performance of female participants was significantly better than males, assessment accuracy of male participants was also at a desired level. Yet, it did not exhibit a noticeable change by age or expertise.

Another important factor regarding evaluation of fabric hand by a sensory panel is the number of panel members. Working with a large panel may provide a more precise representation of consumers' opinion on the total hand value. However, determining the optimum number of participants to minimize the required time and labor is a more reasonable approach. Even though it is a commonly preferred sampling method, selecting assessors according to convenience sampling may limit the ability of the sensory panel to represent the general population. For this reason, effect of panel size on accuracy of sensory evaluation results was investigated for both convenience and random sampling conditions. Results of the study pointed that it is possible to obtain accurate data even with 5 assessors with both sampling methods. In fact, increasing the panel size caused significant decreases in accuracy and inter-rater agreement. Based on these findings, it was concluded that working with small panel sizes may provide more reliable results.

## 4. Conclusion

Sensory tests are fundamentals of a comprehensive fabric hand evaluation. The aim of this study was to investigate the effect of panel size and demographic variables on sensory evaluation results and propose guidelines for future studies. The investigations were carried out with a panel of 200 assessors (100 females and 100 males), between ages 19 and 66, with different levels of expertise.

Correlation measures indicated that 140 of 200 assessors have high assessment accuracy, and the likelihood of selecting an assessor with low evaluation validity was only 5.5%. It was observed that working with a small number of assessors can provide sufficient data, regardless of demographic characteristics of the panel members or the sampling method (convenience or random).

Statistical investigations pointed that assessments of female panel members have a higher accuracy than males. It was also proved that the knowledge of textiles has a significant effect on accuracy, meanwhile the effect of age was less prominent. Highest correlation and concordance coefficients were recorded for adult females and expert females. Based on these findings, it was concluded that including female assessors with a higher level of expertise can significantly increase the accuracy of sensory evaluations.

## Acknowledgements

## Declaration

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author also declared that this article is original, was prepared in accordance with international publication and research ethics, and ethical committee permission or any special permission is not required.

## Author Contributions

N. Uren developed the methodology, performed the sensory evaluations and data analysis, and wrote the manuscript.

## References

1. Ciesielska-Wróbel I.L., and L. Van Langenhove, *The hand of textiles – definitions, achievements, perspectives – a review*, Textile Research Journal, 2012. **82**(14): p. 1457-1468.

2. Liao X., J. Hu, Y. Li, Q. Li, and X. Wu, *A review on fabric smoothness-roughness sensation studies*, Journal of Fiber Bioengineering & Informatics, 2011. **4**(2): p. 105-114.

3. Hu X., Z. Chen, and F. Sun, *Digitization of fabric comfort: a multidimensional evaluation strategy to human perceptions of sensorial, thermal and acoustic comfort in clothing*, International Journal of Clothing Science and Technology, 2023. **35**(1): p. 162-175.

4. Liu L., L. Wei, and F. Sun, *Simultaneous-integrated evaluation of mechanical–thermal sensory attributes of woven fabrics in considering practical wearing states*, Textile Research Journal, 2021. **91**(23-24): p. 2872-2881.

5. Liao X., Y. Li, J. Hu, X. Wu, and Q. Li, *A simultaneous measurement method to characterize touch properties of textile materials*, Fibers and Polymers, 2014. **15**(7): p. 1548-1559.

6. Wang H., T.J. Mahar, and R. Hall, *Prediction of the handle characteristics of lightweight next-to-skin knitted fabrics using a fabric extraction technique*, Journal of the Textile Institute, 2012. **103**(7): p. 691-697.

7. Strazdiene E., and M. Gutauskas, *New method for the objective evaluation of textile hand*, Fibres and Textiles in Eastern Europe, 2005. **2**(50): p. 35-38.

8. Kim J.O., and B.L. Slaten, *Objective evaluation of fabric hand: Part I: Relationships of fabric hand by the extraction method and related physical and surface properties*, Textile Research Journal, 1999. **69**(1): p. 59-67.

9. Kawabata S., and M. Niwa, *Clothing engineering based on objective measurement technology*, International Journal of Clothing Science and Technology, 1998. **10**(3/4): p. 263-272.

10. ISO 6658, *Sensory analysis - Methodology - General guidance*, International Organization for Standardization, 2017.

11. ISO 5495, *Sensory analysis - Methodology - Paired comparison test*, International Organization for Standardization, 2005.

12. ISO 10399, *Sensory analysis - Methodology - Duo-trio test*, International Organization for Standardization, 2017.

13. ISO 4120, *Sensory analysis - Methodology - Triangle test*, International Organization for Standardization, 2021.

14. AATCC EP5, *Evaluation procedure for fabric hand value*, American Association of Textile Chemists and Colorists Committee, 2020.

15. ISO 8586, *Sensory analysis - Selection and training of sensory assessors*, International Organization for Standardization, 2023.

16. Stratton SJ. *Population Research: Convenience Sampling Strategies.* Prehospital and Disaster Medicine, 2021. **36**(4): p. 373-374.

17. Kolb B., *Choosing participants for qualitative research. In: Marketing research*, London: SAGE Publications Ltd., 2008.

18. ASTM E3000-18, *Standard guide for measuring and tracking performance of assessors on a descriptive sensory panel*, American Society for Testing and Materials, 2018.

19. Dhingra R.C., T.J. Mahar, R. Postle, V.B. Gupta, S. Kawabata, M. Niwa, and G.A. Carnaby, *The objective specification of the handle of men's suiting materials: A comparison of fabric handle assessments in India, Australia, Japan and New Zealand*, Indian Journal of Textile Research, 1983. **8**: p. 9-15.

20. Kim H., and G. Winakor, *Fabric hand as perceived by U.S. and Korean males and females.* Clothing and Textile Research Journal, 1996. **14**(2): p. 133-144.

21. Speijers J., J. H. Stanton, G.R. Naylor, P. Ramankutty, and D. Tester, *Skin comfort of base layer wool garments. Part 3: The effect of ethnicity on perceptions of comfort using Chinese and Australian wearers*, Textile Research Journal, 2015. **85**(11): p. 1167-1180.

22. Keefe F.J., M.A. Lumley, A.L. Buffington, J.W. Carson, J.L. Studts, C.L. Edwards, D.J. Macklem, A.K. Aspnes, L. Fox, and D. Steffey, *Changing face of pain: Evolution of pain research in psychosomatic medicine*, Psychosomatic Medicine. 2002. **64**(6): p. 921-938.

23. Bacci L., F. Camilli, S. Drago, M. Magli, E. Vagnoni, A. Mauro, and S. Predieri, *Sensory evaluation and instrumental measurements to determine tactile properties of wool fabrics*, Textile Research Journal, 2012. **82**(14): p. 1430-1441.

24. Xue Z., X. Zeng, L. Koehl, and Y. Chen, *Extracting fabric hand information from visual representations of flared skirts*, Textile Research Journal, 2014. **84**(3): p. 246-266.

25. Suzuki Y., and S. Sukigara, *Mechanical and tactile properties of plain knitted fabrics produced from rayon Vortex yarns*, Textile Research Journal, 2013. **83**(7): p. 740-751.

26. Hui C.L., T.W. Lau, S.F. Ng, and K.C.C. Chan. *Neural network prediction of human psychological perceptions of fabric hand*, Textile Research Journal, 2004. **74**(5): p. 375-383.

27. Temel M., A.A. Johnson, and A.B. Lloyd, *Evaluating the repeatability of friction coefficient measurements and tactile perceptions in skin–textile interactions across body regions*, Tribology Letters, 2022. **2022**(70): 23.

28. Jiao J., X. Hu, Y. Huang, J. Hu, C. Hsing, Z. Lai, C. Wong, and J.H. Xin, *Neuro-perceptive discrimination on fabric tactile stimulation by Electroencephalographic (EEG) spectra*, PLoS One, 2020. **15**(10): e0241378.

29. McGregor B.A., M. Naebe, H. Wang, D. Tester, and J. Rowe, *Relationships between wearer assessment and the instrumental measurement of the handle and prickle of knitted wool fabrics*, Textile Research Journal, 2015. **85**(11): p. 1140-1152.

30. Ryu H.S., and E. Kyung, *Preference and subjective evaluation of washed fabric hand using conjoint analysis*, Textile Research Journal, 2010. **80**(20): p. 2167-2175.

31. Yokura H., and M. Niwa, *Objective hand measurement of nonwoven fabrics used for the top sheets of disposable diapers*, Textile Research Journal, 2003. **73**(8): p. 705-712.

32. Stanton J.H., J. Speijers, G.R.S. Naylor, S. Pieruzzini, J. Beilby, E. Barsden, and A. Clarke, *Skin comfort of base layer knitted garments. Part 1: Description and evaluation of wearer test protocol*, Textile Research Journal, 2014. **84**(13): p. 1385-1399.

33. Wilfling J., G. Havenith, M. Raccuglia, and S. Hodder, *Can you see the feel? The absence of tactile cues in clothing e-commerce impairs consumer decision making*, International Journal of Fashion Design, Technology and Education, 2023. **16**(2): p. 224-233.

34. Boles D.B., and S.M. Givens, *Laterality and sex differences in tactile detection and two-point thresholds modified by body surface area and body fat ratio*, Somatosensory and Motor Research, 2011. **28**(3-4): p. 102–109.

35. Chen C.C., G.K. Essick, D.G. Kelly, M.G. Young, J.M. Nestor, and B. Masse, *Gender-, side-, and site-dependent variations in human perioral spatial resolution*, Archives of Oral Biology, 1995. **40**(6): p. 539-548.

36. Komiyama O., M. Kawara, and A. De Laat, *Ethnic differences regarding tactile and pain thresholds in the trigeminal region*, The Journal of Pain, 2007. **8**(4): p. 363-369.

37. Wohlert A.B., *Tactile perception of spatial stimuli on the lip surface by young and older adults*, Journal of Speech, Language and Hearing Research, 1996. **39**(6): p. 1191-1198.

38. Roh E.K., K.W. Oh, and S.H. Kim, *Effect of raising cycles on mechanical, comfort, and hand properties of artificial suede*, Textile Research Journal, 2014. **84**(18): p. 1995-2005.

39. Asad R.A., W. Yu, Q. Siddiqui, M. Vincent, and Q. Wang, *Subjective evaluations of fabric-evoked prickle using the unidimensional rating scale from different body areas*, Textile Research Journal, 2016. **86**(4): p. 350-364.

40. Harpa R., C. Piroi, I. Cristian, E. Visileanu, and M. Blaga, *Sensory analysis of textiles: case study of an assortment of stretch denim fabrics*, Industria Textila, 2019. **70**(4): p. 358-365.

41. Ye T., L. Cui, C. Zhang, and Y. Qiu, *Electroneurophysiological responses to fabric-skin dynamic contact with different fabrics among different types of people*, Textile Research Journal, 2023. **93**(17-18): p. 4140-4152.

42. Musa A.B.H., B. Malengier, S. Vasile, and L. Van Langenhove, *A comprehensive approach for human hand evaluation of split or large set of fabrics*, Textile Research Journal, 2019. **89**(19-20): p. 4239-4252.

43. Xue Z., X. Zeng, L. Koehl, and L. Shen, *Consistency and reliability of untrained consumers' perceptions of fabric hand of men's suiting*, Textile Research Journal, 2016. **86**(13): p. 1425-1442.

44. Ritchie H., and M. Roser, [cited 2023 23 October] Available from: https://ourworldindata.org/age-structure

45. Hanada M., *Tactile dimensions of fabrics expressed by Japanese onomatopoeic words and phonemic features related to fabric luxuriousness and pleasantness*, Frontiers in Language Sciences, 2023. **2**: 1075055.

46. Wang Q., Y. Tao, Z. Zhang, J. Yuan, Z. Ding, Z. Jiang, Z.

Jia, and J. Wang, *Representations of fabric hand attributes in the cerebral cortices based on the Automated Anatomical Labeling atlas*, Textile Research Journal, 2019. **89**(18): p. 3768-3778.

47. Tang W., S. Zhang, C. Yu, H. Zhu, S. Chen, and Y. Peng, *Tactile perception of textile fabrics based on friction and brain activation*, Friction, 2023. **11**(7): p. 1320-1333.

48. Vasile S., B. Malengier, A. De Raeve, and F. Deruyck, *Influence of selected production parameters on the hand of mattress knitted fabrics assessed by the Fabric Touch Tester*. Textile Research Journal, 2019. **89**(1): p. 98-112.

49. Uren N., and A. Okur, *Analysis and improvement of tactile comfort and low-stress mechanical properties of denim fabrics*, Textile Research Journal, 2019. **89**(23-24): p. 4842-4857.

50. Shao Y., Y. Sun, D. Zheng, G. Liu, Z. Du, J. Liu, and M. Wang, *Tactile comfort characterization of knitted fabrics based on the ring-shaped style tester*, Textile Research Journal, 2021. **91**(7-8): p. 766-777.

51. Sun F., Z. Du, and M. Naebe, *Determination of model parameters for predicting handle characteristics of wool-rich suiting woven fabrics based on the Wool HandleMeter and KES-F*. Journal of the Textile Institute, 2018. **109**(2): p. 147-159.

52. Sun F., Z. Du, D. Zheng, X. Hu, Y. Sun, and W. Gao, *In-situ characterization of handle characteristics of suiting woven fabrics by a simultaneous measurement method*, Textile Research Journal, 2019. **89**(13): p. 2522-2531.

53. Mason M., *Sample size and saturation in PhD studies using qualitative interviews*, Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 2010. **11**(3): 4902282.

54. Nascimento L.C.N., T.V. Souza, I.C.S. Oliveira, J.R.M.M. Moraes, R.C.B. Aguiar, and L.F. Silva, *Theoretical saturation in qualitative research: An experience report in interview with schoolchildren*. Revista Brasileira de Enfermagem, 2018. **71**(1): p. 228-23.