



A HYBRID DECISION SUPPORT SYSTEM APPLICATION WITH THE ANALYTIC HIERARCHY PROCESS AND DATA MINING TECHNIQUES: DIAGNOSIS OF COVID19 WITH COMPLETE BLOOD COUNT VALUES

ANALİTİK HİYERARŞİ SÜRECİ VE VERİ MADENCİLİĞİ TEKNİKLERİYLE HİBRİT BİR KARAR DESTEK SİSTEMİ UYGULAMASI: TAM KAN SAYIMI DEĞERLERİ İLE KOVID19 TANISI

Ahmet Bursalı¹ , Aslı Suner^{2*} 

¹Huawei Turkey R&D Center, Intelligent Application DC, İstanbul, Türkiye

²Department of Biostatistics and Medical Informatics, Faculty of Medicine, Ege University, İzmir, Türkiye

ABSTRACT

Objective: Data mining techniques have a significant impact on enhancing the precision of diagnostics based on artificial intelligence. In this research, it was aimed to develop a web-based decision support that predicts the status of a person who comes to the hospital with Covid-19 suspicion by using complete blood count results until the imaging and PCR test results are obtained.

Method: In this study, firstly data pre-processing techniques on the data set were applied, then feature selection was made using data mining approaches. After reducing the number of variables, the analytical hierarchy process method (AHP), a prominent multi-criteria decision-making approach, was utilized. Through the AHP method combined with expert opinions, the priorities of the variables determined by machine learning were ascertained, leading to the development of a decision model using publicly accessible data. A web-based application of this decision model was subsequently crafted to provide the decision support system to the end-users. Furthermore, an evaluation was conducted to gauge the usability of the decision support system and the satisfaction of its users.

Results: RFE-SVM feature selection algorithm identified seven pivotal variables: Basophil, Eosinophil, Lymphocyte, Leukocyte, Neutrophil, Platelet, and Monocyte. Consultations were held with six expert physicians spanning diverse specialties relevant to COVID-19 diagnosis decision-making with the AHP method. Out of the 42 expert users (57.1% were male, with an average age of 37.30±10.56) were evaluated the system. The System Usability Scale (SUS) score averaged 81.43±15.64, indicating high usability.

Conclusion: Consequently, this system might enable faster isolation of the patient and the commencement of preliminary treatment.

Key Words: Covid-19, Machine Learning, Imbalance Data, Feature Selection, Decision Support System, Analytic Hierarchy Process (AHP) Method

ÖZ

Amaç: Veri madenciliği teknikleri, yapay zeka temelli tanı doğruluğunu artırmada önemli bir etkiye sahiptir. Bu çalışmada, hastaneye Covid-19 şüphesiyle gelen bir kişinin, görüntüleme ve PCR testi sonuçları elde edilene dek, tam kan sayımı sonuçları kullanılarak, Covid-19 olma durumu hakkında tahminde bulunan bir web tabanlı karar desteği geliştirilmesi amaçlanmıştır.

Yöntem: Bu çalışmada öncelikle veri seti üzerinde veri ön işleme teknikleri uygulanmış, daha sonra veri madenciliği yaklaşımları kullanılarak özellik seçimi yapılmıştır. Değişken sayısı azaltıldıktan sonra çok kriterli karar verme yaklaşımının önde gelenlerinden analitik hiyerarşi süreci yöntemi (AHP) kullanılmıştır. Uzman görüşleri ile birleştirilen AHP yöntemiyle makine öğrenmesiyle elde edilen değişkenlerin öncelikleri belirlenmiş ve kamuya açık veriler kullanılarak bir karar modeli geliştirilmiştir. Bu karar modelinin bir web tabanlı uygulaması, daha sonra son kullanıcılara karar destek sistemi sağlamak üzere hazırlanmıştır. Ayrıca, karar destek sisteminin kullanılabilirliğini ve kullanıcı memnuniyetini ölçmek için bir değerlendirme yapılmıştır.

Bulgular: RFE-SVM özellik seçimi algoritması yedi önemli değişkeni tanımlamıştır: Bazofil, Eozinofil, Lenfosit, Lökosit, Nötrofil, Trombosit ve Monosit. AHP yöntemi ile Covid-19 tanısına karar vermeyle ilgili farklı uzmanlık alanlarından altı uzman hekim ile görüşülmüştür. 42 uzman kullanıcı (%57.1'i erkek, yaş ortalaması 37.30±10.56) sistemi değerlendirdi. Sistem Kullanılabilirlik Ölçeği (SUS) puanının ortalaması 81.43±15.64 olup, yüksek kullanılabilirliği göstermektedir.

Sonuç: Sonuç olarak, bu sistem hastanın daha hızlı izole edilmesini ve ilk tedavisinin başlatılmasını sağlayabilir.

Anahtar Kelimeler: Covid-19, Makine Öğrenimi, Dengesiz Veri, Öznitelik Seçimi, Karar Desteği Sistemi, Analitik Hiyerarşi Süreci (AHP) Yöntemi

Makale Bilgisi/Article Info

Yükleme tarihi/Submitted: 10.08.2023, **Revizyon isteği/Revision requested:** 02.10.2023, **Son düzenleme tarihi/Last revision received:** 03.10.2023, **Kabul/Accepted:** 05.10.2023

***Sorumlu yazar/Corresponding author:** Ege University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, İzmir, Türkiye

²Email: asli.suner@ege.edu.tr, ¹Email: ahmetbursalı.deu@gmail.com

**This study was presented orally at the 14th Medical Informatics Congress hosted by İzmir University of Economics between March 16-18, 2023.

INTRODUCTION

COVID-19, caused by the novel SARS CoV-2, first emerged in China and swiftly spread across the globe, posing a significant health threat [1]. From the onset of 2020, the pandemic has deeply impacted human health, global economies, and financial markets [2]. This situation has been especially challenging for countries with limited financial resources dedicated to medical testing and treatment [1]. Diagnosing COVID-19 primarily hinges on a patient's clinical and epidemiological history [3], supplemented by tests such as chest tomography (CT scan) [4]. However, the symptoms associated with COVID-19 are not exclusive, making them insufficient for a definitive diagnosis. While CT scan findings can provide insights, they are not unique to COVID-19 and lack standalone diagnostic value [5]. At present, Real-Time Polymerase Chain Reaction (RT-PCR) tests, which detect viral RNA in samples from the throat or nose, are the gold standard for diagnosing COVID-19 [6]. Yet, RT-PCR has its limitations [7]. Given the pressing need for more accurate tests, global efforts are being made to develop innovative strategies [8]. In this context, the complete blood count (hemogram), a routine blood test, has emerged as a potential diagnostic tool for COVID-19 [9]. Recent studies have delved into using hemogram data to identify potential COVID-19 cases through machine learning techniques and to predict disease outcomes [10]. The digital revolution has facilitated the processing of vast health and molecular datasets, aiming to identify patterns and meaningful correlations, with research in this domain expanding rapidly [11]. The diverse array of machine learning algorithms available necessitates the selection of the most appropriate algorithm tailored to each specific problem.

In this research, our objective was to develop a web-based decision support system capable of accurately predicting SARS-Cov-2 infections using complete blood count values. These values were sourced from a diverse patient group amenable to data sharing. We integrated the Analytic Hierarchy Process (AHP) with data mining techniques to achieve this. While the RT-PCR test remains the predominant diagnostic tool for COVID-19, our study introduces an alternative diagnostic avenue, enabling precise identification of infections based solely on complete blood count values [12]. Following data preprocessing, we deployed machine learning algorithms, streamlined the dataset by eliminating irrelevant variables through feature selection methods [13], and addressed class imbalance issues. We employed the most frequently cited machine learning classification algorithms from existing literature to pinpoint the highest-performing ones. Expert insights were incorporated into our research. To validate the significance of the identified variables, an AHP model was constructed based on expert consensus. Subsequently, an online web application was developed, centered around the most critical variable identified by this model. This tool allows healthcare professionals to make preliminary assessments about a patient's COVID-19 status using their complete blood count, even before RT-PCR test or CT scan results are available. Consequently, potential COVID-19 patients can be isolated promptly, facilitating timely initiation of treatment.

METHOD

Outlined in Figure 1, the study's overarching workflow begins with an introduction to the utilized dataset. Following data preprocessing, we addressed class imbalance issues and proceeded with feature selection to identify pivotal variables. The commonly used methods to eliminate class imbalance including random oversampling, random undersampling, smote, smote-tomeklink, adasyn algorithms were applied. The performance metrics of accuracy, sensitivity, specificity, precision and F1 score were used to compare the classification performances of the machine learning algorithms, and the execution times of these algorithms were documented.

The performance metrics derived from the analyses were segmented into four primary categories: (i) Evaluation using the original dataset, (ii) Evaluation post-variable selection, (iii) Analysis after balancing the

imbalanced dataset, and (iv) Evaluation of the balanced dataset after variable selection.

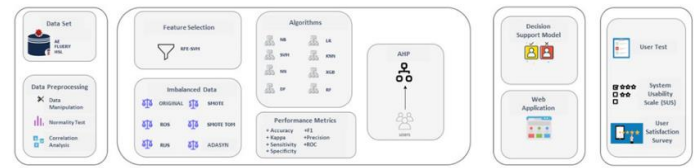


Figure 1. Workflow of the study.

To prioritize variables identified by feature selection methods, we incorporated expert opinions using the AHP formulated by Saaty in the 1970s [14]. The AHP method, tailored to the decision-maker's personal objectives, encompasses three pivotal stages: establishing a hierarchy, analyzing data priorities, and validating consistency [15]. The web application, developed using R-Shiny, was built upon the variables derived from our analyses. In testing phase of the system, we gauged its usability, user expectations, and satisfaction levels through an online survey. This survey incorporated questions crafted by our team and concluded with the "System Usability Scale (SUS)", which was adapted to Turkish in 2011, offering a subjective assessment of the software's efficacy [16].

Statistical Analysis

For user data, categorical variables were represented in frequency tables, while continuous variables were summarized using descriptive statistics. Relationships between categorical variables were explored using the chi-square test. The Shapiro-Wilk normality test assessed the distribution of numeric data. Given the non-normal distribution of the data, we employed the Mann-Whitney U test for two groups and the Kruskal-Wallis test for comparisons involving more than two independent groups. A p-value < 0.05 was considered statistically significant. Statistical analyses were performed using IBM SPSS version 25.0 statistical software.

RESULTS

Table 1 shows a comprehensive overview of the datasets utilized in this study. These datasets, sourced from an open COVID-19 cases database in Brazil [1] include contributions from three prominent institutions: Fleury Group (<https://www.fleury.com.br>), Albert Einstein Hospital (<https://www.einstein.br>), and Sírío Libanês Hospital (<https://www.hospitalsiriolibanes.org.br>). The data encompasses patient records from February 26, 2020 to June 30, 2020 for COVID-19 cases and control data (for individuals not diagnosed with COVID-19) spanning from November 1, 2019 to June 30, 2020. It is noteworthy that all shared patient data have been anonymized, adhering to the highest international standards and guidelines. Within the dataset, only the gender attribute is categorical, with the remainder being numerical. The datasets are devoid of any missing observations. Given the analogous distributions across the three datasets, they were amalgamated into a singular dataset. This merging aimed to facilitate more robust training on a larger, class-balanced dataset.

Table 1. Overview of the datasets used in the study

Data set	AE	FLE	HSL	COVID19
Sample size	4567	803	515	5885
Number of features	22	22	22	22
Number of class	2	2	2	2
Number of categorical feature	1	1	1	1
Number of continues feature	21	21	21	21
Percentage of missing data	0	0	0	0
Ratio of class imbalance	2.262	2.136	41.916	2.242

The dataset encompasses 20 complete blood count variables, namely: Hematocrit (%), Hemoglobin (g/dl), Platelets ($\times 10^3 \mu\text{l}$), Mean Platelet Volume (MPV) (fl), Red Blood Cells (RBC) ($\times 10^6 \mu\text{l}$), Lymphocytes ($\times 10^3 \mu\text{l}$), Leukocytes ($\times 10^3 \mu\text{l}$), Basophils ($\times 10^3 \mu\text{l}$), Eosinophils ($\times 10^3 \mu\text{l}$), Monocytes ($\times 10^3 \mu\text{l}$), Neutrophils ($\times 10^3 \mu\text{l}$), Mean Corpuscular Volume (MCV) (fl), Mean Corpuscular Hemoglobin (MCH) (pg), Mean Corpuscular Hemoglobin Concentration (MCHC) (g/dl), Red Blood Cell Distribution Width (RBDW) (%), %Basophils, %Eosinophils, %Lymphocytes, %Monocytes, %Neutrophils. From these, the RFE-SVM feature selection algorithm identified seven pivotal variables: Basophil, Eosinophil, Lymphocyte, Leukocyte, Neutrophil, Platelet, and Monocyte.

In terms of computational efficiency, it was observed that training durations generally exceeded testing durations. The XGB algorithm was notably the most time-intensive, with training and testing durations of 10.705 minutes and 7.239 minutes, respectively; when all variables from the original data were considered. In contrast, the LR algorithm showcased the swiftest performance, clocking in at 0.096 minutes for training and 0.080 minutes for testing on the same data subset. A significant reduction in runtime was evident post the feature selection process. For a comprehensive breakdown of the performance metrics derived from the various methodologies and algorithms employed in this study, refer to Table 2 below.

Evaluation Using the Original Dataset

Upon analyzing the original dataset, all classification models yielded accuracy scores between 0.751 and 0.782. The Kappa value fluctuated from 0.361 to 0.514, with the RF algorithm achieving the peak accuracy (0.782) and Kappa value (0.514). F1 scores spanned from 0.618 to 0.683. Notably, the NB and LR models demonstrated superior sensitivity compared to their counterparts, making them viable options for detecting true positive cases. However, they also exhibited the lowest specificity, suggesting potential bias. The RF and SVM algorithms displayed balanced sensitivity and specificity values. Overall, SVM-Radial and RF emerged as top performers for the original data.

Analysis with Feature Selection

In this approach, classification methods were compared on the original data after performing feature selection using the RFE-SVM method, as referenced by [13]. This reduced the variables from 20 to 7. The NN algorithm achieved the highest accuracy (0.695-0.778) and Kappa value (0.381-0.511). Notably, the NB and LR models demonstrated superior sensitivity, making them viable options for detecting true positive cases. Overall, standout models varied across algorithms.

Balancing the Imbalanced Dataset and Analyzing

For this approach, classification metrics were derived using five distinct imbalance data processing techniques on the original dataset:

ROS Algorithm: RF achieved the highest accuracy (0.707-0.783), while SVM-Radial led in Kappa value (0.412-0.538). Despite RF's general robustness, it showed reduced sensitivity in the ROS method, possibly due to the ROS's sample generation approach. However, RF excelled in specificity.

RUS Algorithm: RF and SVM-Radial dominated in accuracy scores (0.676-0.778). SVM-Radial had the top Kappa value, while SVM-Linear excelled in sensitivity. RF stood out in specificity.

SMOTE Algorithm: RF led in accuracy (0.675-0.775), while SVM-Radial had the highest Kappa value. SVM-Linear was the top performer in sensitivity, and RF led in specificity.

SMOTE-TOMEKLINK Approach: RF again led in accuracy (0.696-0.780) and Kappa value, LR showcased the highest sensitivity, and DT led in specificity.

ADASYN Algorithm: RF achieved the highest accuracy (0.719-0.779), SVM-Linear led in Kappa and sensitivity, while RF dominated in specificity.

Generally, RF and SVM-Radial consistently emerged as the standout models across various data balancing techniques.

Analysis on Balanced Dataset with Feature Selection

This approach examined the impact of classifier algorithms after applying imbalanced data processing techniques to the dataset post-feature selection. The variable count was again reduced from 20 to 7. Performance metrics were derived using five distinct imbalance data processing techniques:

ROS Algorithm: RF achieved the highest accuracy (0.742-0.775), while KNN led in Kappa value (0.449-0.528). RF's sensitivity was notably lower, but it excelled in specificity. SVM-Radial and RF emerged as top performers.

RUS Algorithm: SVM-Radial dominated in accuracy (0.679-0.776) and Kappa value. SVM-Linear showcased superior sensitivity, while XGB led in specificity, SVM models and RF stood out.

SMOTE Algorithm: KNN led in accuracy (0.616-0.777) and Kappa values. SVM-Radial had the highest sensitivity, and RF achieved the best specificity. SVM-Radial and RF were the standout models.

SMOTE-TOMEKLINK Approach: KNN achieved the highest accuracy (0.702-0.777) and Kappa values. LR excelled in sensitivity, while XGB had the top specificity.

ADASYN Algorithm: KNN led in accuracy (0.718-0.771) and Kappa values. LR showcased the best sensitivity, and RF dominated in specificity. LR, RF, and KNN emerged as the standout models.

In conclusion, while feature selection didn't drastically alter metrics, the ability to achieve better results with fewer variables offers significant advantages in cost and time efficiency.

AHP Method Results

In this research, the AHP method was employed to prioritize criteria for variables derived post-feature selection. For this purpose, consultations were held with six expert physicians spanning diverse specialties relevant to COVID-19 diagnosis decision-making. These specialties included public health, infectious diseases, emergency medicine, chest diseases, internal medicine, and radiology (refer to Figure 2). The experts' average age was 55.83 with a standard deviation of 6.55, ranging from 45 to 63 years.

Development of the Decision Support Model and Web Application

A web-based decision support model was developed, accessible via: <http://185.106.208.185:3838> (as seen in Figure 3). Within this application, users can input blood test results for an individual. Upon clicking the "calculate" button, the system provides a predictive result and probability score for the individual's potential COVID-19 diagnosis. For user convenience, sample data entries for both COVID-19 positive and negative results are available for testing.

Figure 2 showcases steps from a sample application screenshot. In Figure 2a, data for a patient with a 59% likelihood of having COVID-19 is displayed. Conversely, Figure 2b presents a patient with a 73% probability of being healthy. The "Graphs" menu offers visual insights: users can view pie and bar charts based on the results after relevant analyses (as seen in Figure 2c). The pie chart delineates the probabilities of being COVID-negative or positive, while the bar chart illustrates the patient's specific blood result levels.

Experts who contributed to the AHP method with their expert opinions also used the system and expressed their opinions in the user survey.

Table 2. Detailed performance metrics for the implemented methods and algorithms

Method	Algorithm	Accuracy		Kappa		Sensitivity		Specificity		Precision		F1-Score		ROC	
		ALL	FS	ALL	FS	ALL	FS	ALL	FS	ALL	FS	ALL	FS	ALL	FS
ORIGINAL	NB	0.751	0.757	0.477	0.478	0.699	0.646	0.782	0.825	0.664	0.694	0.681	0.669	0.806	0.799
	SVM_L	0.778	0.771	0.508	0.492	0.594	0.585	0.892	0.886	0.772	0.759	0.671	0.661	0.818	0.814
	SVM_R	0.781	0.772	0.513	0.491	0.592	0.568	0.897	0.898	0.779	0.773	0.673	0.655	0.842	0.830
	LR	0.683	0.695	0.361	0.381	0.722	0.713	0.659	0.684	0.565	0.581	0.634	0.640	0.740	0.741
	KNN	0.777	0.776	0.501	0.502	0.576	0.591	0.900	0.889	0.780	0.766	0.663	0.667	0.835	0.828
	RF	0.782	0.768	0.514	0.484	0.588	0.570	0.901	0.890	0.785	0.761	0.672	0.652	0.839	0.823
	DT	0.753	0.748	0.444	0.430	0.525	0.509	0.893	0.895	0.751	0.748	0.618	0.606	0.758	0.756
	XGB	0.760	0.752	0.460	0.436	0.536	0.496	0.898	0.910	0.763	0.773	0.629	0.604	0.825	0.816
	NN	0.775	0.778	0.510	0.511	0.635	0.607	0.861	0.884	0.737	0.763	0.683	0.676	0.816	0.819
ROS	NB	0.744	0.747	0.474	0.470	0.743	0.698	0.746	0.778	0.642	0.659	0.689	0.678	0.804	0.795
	SVM_L	0.760	0.754	0.508	0.496	0.771	0.772	0.754	0.742	0.658	0.648	0.710	0.705	0.818	0.812
	SVM_R	0.777	0.772	0.538	0.527	0.763	0.756	0.786	0.782	0.687	0.681	0.723	0.717	0.839	0.830
	LR	0.707	0.742	0.412	0.474	0.765	0.762	0.671	0.730	0.588	0.634	0.665	0.692	0.775	0.806
	KNN	0.760	0.774	0.504	0.528	0.753	0.743	0.764	0.793	0.662	0.688	0.705	0.714	0.833	0.828
	RF	0.783	0.775	0.524	0.506	0.631	0.621	0.876	0.869	0.758	0.745	0.689	0.677	0.839	0.822
	DT	0.743	0.743	0.462	0.449	0.695	0.632	0.773	0.812	0.653	0.674	0.673	0.652	0.770	0.748
	XGB	0.768	0.763	0.509	0.497	0.705	0.692	0.806	0.806	0.691	0.687	0.698	0.689	0.824	0.815
	NN	0.752	0.772	0.471	0.522	0.661	0.729	0.808	0.799	0.679	0.690	0.670	0.709	0.830	0.698
RUS	NB	0.750	0.754	0.485	0.485	0.746	0.708	0.753	0.782	0.650	0.667	0.694	0.687	0.804	0.799
	SVM_L	0.759	0.752	0.507	0.495	0.775	0.780	0.749	0.736	0.655	0.645	0.710	0.706	0.819	0.812
	SVM_R	0.778	0.776	0.536	0.532	0.746	0.743	0.798	0.797	0.694	0.692	0.719	0.716	0.838	0.829
	LR	0.676	0.679	0.357	0.360	0.754	0.747	0.628	0.637	0.555	0.558	0.639	0.639	0.745	0.751
	KNN	0.759	0.772	0.504	0.522	0.760	0.737	0.758	0.793	0.659	0.687	0.706	0.711	0.830	0.824
	RF	0.778	0.769	0.532	0.518	0.723	0.734	0.812	0.791	0.702	0.684	0.713	0.708	0.839	0.820
	DT	0.746	0.741	0.470	0.450	0.713	0.658	0.766	0.792	0.652	0.661	0.681	0.659	0.771	0.748
	XGB	0.765	0.767	0.503	0.503	0.696	0.679	0.808	0.822	0.690	0.700	0.693	0.689	0.823	0.821
	NN	0.774	0.768	0.527	0.510	0.738	0.705	0.796	0.807	0.690	0.692	0.713	0.699	0.826	0.819
SMOTE	NB	0.743	0.751	0.471	0.479	0.738	0.705	0.747	0.780	0.642	0.663	0.687	0.683	0.805	0.796
	SVM_L	0.764	0.616	0.516	0.007	0.775	0.027	0.758	0.979	0.663	0.439	0.715	0.05	0.818	0.696
	SVM_R	0.773	0.771	0.529	0.523	0.757	0.753	0.783	0.781	0.682	0.679	0.718	0.714	0.837	0.829
	LR	0.675	0.664	0.355	0.333	0.754	0.737	0.626	0.619	0.553	0.543	0.639	0.625	0.744	0.733
	KNN	0.750	0.777	0.489	0.536	0.768	0.751	0.739	0.793	0.644	0.691	0.701	0.720	0.832	0.827
	RF	0.775	0.770	0.517	0.506	0.674	0.665	0.837	0.834	0.718	0.712	0.695	0.688	0.834	0.819
	DT	0.733	0.750	0.444	0.462	0.696	0.638	0.756	0.818	0.637	0.683	0.665	0.660	0.771	0.749
	XGB	0.762	0.762	0.496	0.491	0.690	0.664	0.806	0.823	0.686	0.697	0.688	0.680	0.823	0.808
	NN	0.757	0.773	0.502	0.525	0.772	0.734	0.747	0.798	0.653	0.690	0.708	0.711	0.821	0.711
SMOTE-TOMEKLINK	NB	0.755	0.758	0.482	0.477	0.690	0.637	0.794	0.832	0.673	0.699	0.682	0.667	0.804	0.799
	SVM_L	0.778	0.770	0.508	0.490	0.600	0.586	0.887	0.883	0.766	0.755	0.673	0.660	0.818	0.814
	SVM_R	0.778	0.772	0.506	0.491	0.588	0.570	0.895	0.897	0.775	0.772	0.668	0.656	0.841	0.830
	LR	0.696	0.702	0.385	0.390	0.723	0.704	0.680	0.701	0.581	0.591	0.645	0.643	0.748	0.749
	KNN	0.776	0.777	0.498	0.504	0.570	0.582	0.902	0.898	0.782	0.777	0.659	0.666	0.835	0.827
	RF	0.780	0.768	0.513	0.484	0.598	0.576	0.892	0.886	0.773	0.756	0.674	0.654	0.833	0.824
	DT	0.752	0.758	0.437	0.454	0.504	0.525	0.904	0.901	0.764	0.766	0.608	0.623	0.759	0.760
	XGB	0.757	0.750	0.453	0.427	0.533	0.479	0.895	0.916	0.757	0.778	0.625	0.593	0.827	0.813
	NN	0.770	0.767	0.490	0.479	0.586	0.598	0.883	0.871	0.755	0.740	0.660	0.662	0.821	0.819
ADASYN	NB	0.743	0.742	0.477	0.469	0.769	0.740	0.727	0.744	0.634	0.640	0.695	0.686	0.804	0.798
	SVM_L	0.767	0.738	0.524	0.471	0.789	0.784	0.754	0.710	0.663	0.624	0.721	0.695	0.819	0.810
	SVM_R	0.779	0.763	0.544	0.512	0.781	0.768	0.778	0.759	0.684	0.662	0.729	0.711	0.833	0.827
	LR	0.767	0.735	0.522	0.467	0.780	0.789	0.759	0.703	0.666	0.620	0.718	0.694	0.821	0.806
	KNN	0.753	0.771	0.496	0.527	0.778	0.765	0.737	0.775	0.646	0.676	0.706	0.718	0.827	0.823
	RF	0.776	0.764	0.527	0.500	0.714	0.689	0.814	0.811	0.703	0.691	0.708	0.690	0.835	0.813
	DT	0.719	0.741	0.437	0.450	0.783	0.658	0.680	0.792	0.600	0.661	0.680	0.659	0.743	0.748
	XGB	0.747	0.755	0.483	0.485	0.765	0.702	0.737	0.787	0.641	0.670	0.697	0.686	0.816	0.812
	NN	0.763	0.718	0.513	0.431	0.768	0.763	0.760	0.690	0.663	0.602	0.712	0.673	0.824	0.617

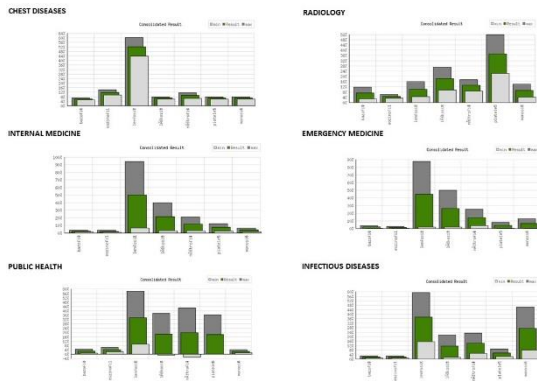


Figure 2. Criterion priorities obtained as a result of the AHP analysis.

User Test Results

The online survey form for user evaluation can be accessed at this link <https://form.jotform.com/223380903422954> (as seen in Figure 4). Out of the 42 expert users who evaluated the system, 57.1% were male, with an average age of 37.30±10.56. A significant majority, 88.1%, are proficient in computer usage. Their educational backgrounds vary: 31.0% in Medicine, 21.4% in Biology/Molecular Biology/Biochemistry, 19.0% in Statistics/Mathematics, and 16.7% in Computer Science/Engineering/Programming. The System Usability Scale (SUS) score averaged 81.43±15.64, indicating high usability. There were no significant differences in SUS scores when segmented by gender, computer proficiency, or academic background (p-value=0.277; p-value=0.714; and p-value=0.731, respectively). Furthermore, 83.4% of users expressed satisfaction with the application, and 76.2% appreciated its design. These sentiments did not vary significantly based on gender, computer skills, or academic backgrounds (p-value=0.684; p-value=0.431; p-value=0.292; p-value=0.432; p-value=0.940; and p-value=0.738, respectively).

Users highlighted several positive aspects of the application: (i) The interface is user-friendly, quick, and straightforward; (ii) The cost-free nature of the application was appreciated; (iii) Users found the probability values in the results particularly useful; (iv) The application aids in determining patient triage priorities; (v) It can assist healthcare professionals in managing patients while awaiting test results; (vi) Knowing the calculated probability of a positive test result can help reduce transmission risks; (vii) Users valued the inclusion of quantitative data through mathematical modeling. However, there were some criticisms: (i) Technical terms like leukocytes and neutrophils require explanatory texts; (ii) A directly downloadable mobile app would enhance user experience; (iii) The introductory directive could be more concise. Survey participants also offered suggestions for improvement: (i) Incorporate additional variables such as age, gender, race, oxygen saturation, blood pressure, and pulse; (ii) Differentiate between patients with and without chronic diseases; (iii) Integrate with e-government systems to send results directly to patients' phones; (iv) Save time by automatically retrieving laboratory findings through integration with hospital information systems; (v) Enhance graphics with added explanations and improved visuals.

DISCUSSION

Machine learning techniques have become instrumental in medical diagnostics, offering insights into disease probabilities and risk factors based on clinical and laboratory data. These methods not only enhance diagnostic accuracy but also alleviate the workload. While data engineering techniques, including data mining and machine learning, have found success in various real-world scenarios, medical diagnostics often demand meticulous data preprocessing. Challenges like insufficient data, cognitive errors from data collection or verification discrepancies, missing observations, class imbalances, and irrelevant variables can skew diagnostic outcomes. Addressing these issues through data preprocessing before deploying data mining

algorithms can significantly boost classification accuracy and overall algorithm performance.

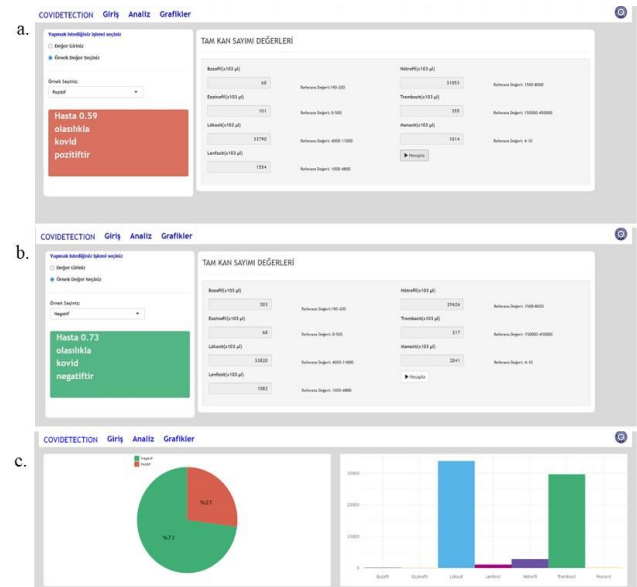


Figure 3. Screenshot of the web-based decision support model

In healthcare, class imbalances can undermine the efficacy of models trained on datasets with underrepresented classes. This imbalance has been addressed in various biomedical contexts, such as breast cancer diagnosis, gene expression data in cancer microarrays, and protein intracellular location data [17]. When considering the challenge of irrelevant variables, some studies have employed embedded feature extraction methods, as seen in research on drug response heterogeneity in Type 2 diabetes patients. In contrast, others have utilized statistical filtering methods, as in thyroid nodule identification studies [18]. Another notable research combined feature selection with cluster analysis techniques for clinical breast cancer diagnosis [19]. A study by [20] focused on developing a decision support system for predicting kidney stone types using ensemble learning. Additionally, a comprehensive study has compared the outcomes of various feature selection methods applied to clinical datasets [21].

The AHP, a subset of Multicriteria Decision Analysis (MCDA) methods, has gained global recognition across diverse sectors, including health, industry, marketing, finance, and more. Recent findings highlight its efficacy in healthcare for diagnosis, treatment prioritization, health management, and health technology assessment [22]. Prior research has emphasized the potential of AHP combined with decision tree methodologies in aiding general surgeons with decision-making in rectal cancer treatments [23,24]. Given the multifaceted nature of healthcare decision-making, AHP, a prominent MCDA method, is a preferred choice [22].

Comparing our findings with the reference study [1], a notable distinction lies in the analytical approach. While the reference utilized Python for analysis, this study employed the R programming language. Differences in utilized libraries and function parameters can influence outcomes. For instance, while the reference leveraged the scikit-learn package in Python, this study used the caret package in R. The reference ran classification algorithms 31 times, recording all results, whereas this study adopted the 10-fold cross-validation method, aiming for more consistent and trustworthy outcomes. Additionally, while the reference article analyzed three separate datasets, this study amalgamated them into one, reducing data imbalance and enhancing training quality. Overall, despite the close alignment of results between the reference and this study, our approach yielded superior outcomes in certain performance metrics.

In a study conducted on the Fleury dataset, [25] aimed to predict Covid-19 outcomes using hemogram results and age through the

Xgboost algorithm. However, the study did not address the class imbalance issue, nor did it undertake a feature selection process. Examining the Xgboost algorithm's classification performance, the accuracy was 0.80, Precision 0.756, Specificity 0.82, F1-score 0.701, and the ROC value was 0.811. In contrast, our study rectified the class imbalance and implemented feature selection, influencing the results.

Another notable study by [13] assessed the impact of feature selection (hemogram values) on Covid-19. This study utilized the Albert Einstein hospital dataset and employed the Recursive Feature Elimination method of SVM for variable selection, similar to our approach. However, their focus was primarily on variable selection, measuring the impact of these variables through statistical analyses. In our study, we gauged the influence of selected variables using machine learning algorithms.

Figure 4. Screenshot of the online survey form for user evaluation

Decision Support Systems (DSS) have gained traction in the medical field, aiding physicians in the diagnostic process. These systems enhance decision-making by offering reminders, alerts, suggestions, and interpretations [26]. Patient-specific characteristics can be input into the DSS, either by the physician, patient, or through electronic medical records, to generate tailored recommendations. In our study, we developed an online decision support model that predicts the likelihood of a patient having Covid-19. This facilitates quicker isolation and preliminary treatment for suspected Covid-19 patients

while awaiting further test results. Our study stands out for its unique approach in creating a web-based decision support system using data mining with hemogram data for Covid-19 diagnosis.

Limitations

One limitation of our study is its reliance on datasets from three specific hospitals in Brazil, without incorporating data from other countries or hospitals. Furthermore, the developed application is recommendation-based, intended to complement, not replace, primary diagnostic methods for detecting Covid-19, such as PCR and Tomography.

CONCLUSION

This study underscores the significance of hemogram data in diagnosing Covid-19. While it's designed to complement existing diagnostic methods, the rapid availability of full blood count test results can provide preliminary insights into Covid-19, facilitating quicker decisions. As comprehensive research encompassing diverse datasets and methodologies, it offers valuable insights for future research. The developed web application allows researchers of varying expertise to create and utilize explainable models for predictions, providing a user-friendly, free platform with reproducible outcomes.

Ethical Approval: Not necessary.

Conflict of Interest: The authors have no conflicts of interest to declare.

Funding: This study was supported by the Ege University Scientific Research Projects Coordination Unit (BAP) (Project Code: TGA-2021-23066).

Acknowledgements: The authors would like to express their sincere gratitude to the experts who shared their expert opinions with them in this study and to the users who evaluated the decision support application for their contributions.

Author Contribution: Concept: AS; Design: AS; Data collecting: AS, AB; Statistical analysis: AB; Literature review: AS, AB; Writing: AS, AB; Critical review: AS.

REFERENCES

- Dorn M, Grisci BI, Narloch PH, et al. Comparison of machine learning techniques to handle imbalanced COVID-19 CBC datasets. PeerJ Comput Sci. 2021;7:1-34.
- Nicola M, Alsafi Z, Sohrabi C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg. 2020;78(3):185-193.
- Ge H, Wang X, Yuan X, et al. The epidemiology and clinical information about COVID-19. Eur J Clin Microbiol Infect Dis. 2020;39(6):1011-1019.
- Bernheim A, Mei X, Huang M, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology. 2020;295(3):200463.
- Hope MD, Raptis CA, Shah A, Hammer MM, Henry TS. A role for CT in COVID-19? What data really tell us so far. Lancet. 2020;395(10231):1189-1190.
- Hadaya J, Schumm M, Livingston EH. Testing Individuals for Coronavirus Disease 2019 (COVID-19). JAMA. 2019;2020.
- Vogels CBF, Brito AF, Wyllie AL, et al. Grubaugh ND. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. Nat Microbiol. 2020;5(10):1299-1305.
- Zame WR, Bica I, Shen C, et al. M. Machine learning for clinical trials in the era of COVID-19. Stat Biopharm Res. 2020;12(4):506-517.
- Ferrari D, Motta A, Strollo M, Banfi G, Locatelli M. Routine blood tests as a potential diagnostic tool for COVID-19. Clin Chem Lab Med. 2020;58(7):1095-1099.
- Yan L, Zhang HT, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell. 2020;2(5):283-288.
- Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. Informatics in Medicine Unlocked. 2021;24:100564.
- Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395(10229):1054-1062.
- Feltes BC, Vieira IA, Parraga-Alava J, et al. Feature selection reveal peripheral blood parameter's changes between COVID-19 infections patients from Brazil and Ecuador. Infect Genet Evol. 2022;98:105228.

14. Saaty RW. The analytic hierarchy process-what it is and how it is used. *Math Modelling*. 1987;9(3-5):161-176.
15. Ho W. Integrated analytic hierarchy process and its applications - A literature review. *Eur J Oper Res*. 2008;186(1):211-228.
16. Brooke J. SUS-A quick and dirty usability scale. *Usability Eval Ind*. 1996;189-194.
17. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell*. 2016;5(4):221-232.
18. Bibicu D, Moraru L, Biswas A. Thyroid nodule recognition based on feature selection and pixel classification methods. *J Digit Imaging*. 2013;26(1):119-128.
19. Chen CH. A hybrid intelligent model of analyzing clinical breast cancer data using clustering techniques with feature selection. *Appl Soft Comput J*. 2014;20:4-14.
20. Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. *Artif Intell Med*. 2018;84:117-126.
21. Chen CW, Tsai YH, Chang FR, Lin WC. Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. *Expert Syst*. 2020;37(5):1-10.
22. Adunlin G, Diaby V, Xiao H. Application of multicriteria decision analysis in health care: A systematic review and bibliometric analysis. *Health Expect*. 2015;18(6):1894-1905.
23. Suner A, Karakulah G, Dicle O, Sökmen S, Çelikoğlu CC. CorrecTreatment: A web-based decision support tool for rectal cancer treatment that uses the analytic hierarchy process and decision tree. *Appl Clin Inform*. 2015;6(1):56-74.
24. Suner A, Çelikoğlu CC, Dicle O, Sökmen S. Sequential decision tree using the analytic hierarchy process for decision support in rectal cancer. *Artif Intell Med*. 2012;56(1):59-68.
25. Silveira EC. Prediction of COVID-19 from hemogram results and age using machine learning. *Front Health Inform*. 2020;9(1):39.
26. Wyatt JC. Decision support systems. *J R Soc Med*. 2000;93(12):629-633.

Karya Journal of Health Science is licensed by [Creative Commons Attribution-NonCommercial-No Derivative 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

