

Research Article

Detection of Mental Well-Being Status Through Data-Driven Approaches

Aysun Kazak Saltı ^{1,*}  and Evrim Ersin Kangal ² 

Received: 13.01.2025

Accepted: 25.02.2025

¹Mersin University, Vocational School of Health Sciences, Medical Services and Techniques Department, First and Emergency Aid, Mersin, Türkiye; aysn1108@gmail.com

²Computer Technology and Information Systems, School of Applied Technology and Management of Erdemli, Mersin University, Erdemli/Mersin, Türkiye; evrimersin@gmail.com

*Corresponding Author.

Abstract: Mental well-being disorders are among the most significant challenges in the modern lifestyles, and it is well-established that early detection of diseases is essential for effective prevention. On the other hand, Machine Learning (ML) algorithms currently play a valuable role in disease detection. The aim of this study is not only to investigate the performance of various ML classifiers but also to propose a modern technique for mental health diagnosis. In this context, our research considers Bootstrap Aggregating (Bagging), Extremely Randomized Trees (ExtraTrees), Passive-Aggressive, Light Gradient Boosting Machine (LGBM), Perceptron, and Stochastic Gradient Descent (SGD) algorithms, which are among the widely recognized ML classifiers in literature. To address the factors contributing to mental health illnesses among the selected individuals, we employ a three-phase data processing approach: segmentation, feature extraction, and classification. Analyzing feature importance from the selected dataset, our study highlights the significant impact of age, family history, and workplace environment on a worker's mental health status.

Keywords: machine learning; classifier; health

Araştırma Makalesi

Veri Odaklı Yaklaşımlar ile Zihinsel İyi Oluş Durumunun Tespiti

Özet: Ruhsal iyi oluş düzeyi bozuklukları, modern yaşam biçimlerinin en önemli sorunları arasında yer almaktadır ve hastalıkların erken teşhisinin etkili bir önleme için ne kadar önemli olduğu artık bilinen bir gerçektir. Diğer yandan, Makine Öğrenimi (MÖ) algoritmaları günümüzde hastalık tespiti için değerli bir rol oynamaktadır. Bu çalışmanın amacı yalnızca çeşitli ML sınıflandırıcılarının performansını araştırmak değil aynı zamanda ruh sağlığı tanısı için modern bir teknik sunmaktır. Bu bağlamda, araştırmamızda literatürde yaygın olarak bilinen ML sınıflandırıcılarından Bootstrap Aggregating (Bagging), Extremely Randomized Trees (ExtraTrees), Passive-Aggressive, Light Gradient Boosting Machine (LGBM), Perceptron ve Stochastic Gradient Descent (SGD) algoritmaları ele alınmıştır. Seçilen bireylerde ruhsal sağlık hastalıklarına katkıda bulunan faktörleri ele almak için üç aşamalı bir veri işleme yaklaşımı kullanıyoruz: parçalara ayırma, özellik çıkarımı ve sınıflandırma. Ele alınan veri setindeki özellik önemini analiz eden araştırmamız, yaşın, aile geçmişinin ve işyeri ortamının bir çalışanın ruh sağlığı durumu üzerinde önemli bir etkiye sahip olduğunu ortaya koymaktadır.

Anahtar Kelimeler: makine öğrenme; sınıflandırıcı; sağlık

1. Introduction

Data science relies on advanced methodologies and state-of-the-art numerical analysis algorithms to derive insights from data. In addition to this, the ML is a branch of Artificial Intelligence (AI) that enables computers to learn from large datasets through training without requiring explicit programming [1,2]. A learning machine refers to a high-dimensional input–output mapping, expressed as

$$\underset{\text{Algorithm}}{f} \left(\underset{\text{input}}{x} \right) = \underset{\text{Output}}{y}, \quad (1.1)$$

where the output may be discrete, continuous, or a combination of both. These algorithms process data, extract insights, and make determinations or predictions. Furthermore, classification is a data mining approach aimed at predicting the specific group memberships among diverse data instances and there is a variety of techniques tailored for this purpose [3,4]. A significant milestone in the ML domain was the development of Support Vector Machines (SVM), introduced by Vapnik and Cortes in 1995 [5], renowned for their strong theoretical foundation and precise results. Given the pivotal role of ML algorithms in analyzing data obtained from specific devices or tools, researchers across various disciplines such as medicine [6-8], economy [9, 10], cosmology [11-13], astrophysics [14], and meteorology [15, 16] are increasingly employing ML approaches to streamline and enhance their investigations.

On the other hand, the prevalence of mental illness is escalating rapidly on a global scale and the World Health Organization (WHO) projects that one in four individuals will encounter mental or neurological disorders at some point in their lives [17]. Mental health diagnosis is a complex process that involves multiple steps and is not straightforward. It typically begins with a carefully structured interview that includes questions regarding the symptoms and medical history, and may also involve a physical examination in some cases [18, 19]. Diverse psychological assessments are also administered to ensure that the symptoms are solely attributable to the mental health condition and not to any other underlying issues. The overlap of factors and symptoms across various mental health conditions makes health condition diagnoses increasingly complex and, in some cases, may lead to the possibility of misdiagnosis. As the ML methods, grounded in concepts from probability and other disciplines, are well-suited to manage uncertain or incomplete information, researchers and healthcare professionals have progressively utilized the ML techniques for the analysis of mental health data. For instance, Reddy et al. [20] utilized six ML techniques to investigate stress patterns in working adults and to identify the primary factors that strongly influence stress levels. Next, in Ref. [21], the authors analyzed a dataset collected for the mental state of employees in a workplace and examined the factors influencing their mental health status via various ML approaches. Jain et al. [22] used eight well-known ML algorithms for a dataset of 1,429 individuals to construct precise and reliable predictive models, with the aim of improving decision-making accuracy and contributing to the reduction of suicide cases through diverse methodological approaches. Garriga et al. [23] proposed an ML model that utilizes electronic health records to monitor patients continuously for the risk of a mental health crisis within a 28-day timeframe. Rezapour and Hansen [24] focused on a survey data from the Inter-university Consortium for Political and Social Research and considered a range of statistical and ML techniques to investigate the effects of the COVID-19 pandemic on the mental health of frontline workers in the United States.

This investigation is based on a comprehensive mental health dataset [25] of 1,257 patients, encompassing a broad spectrum of variables, including demographic attributes, care options, family history, environmental conditions, workplace quality, and health-related behaviors. The variation in characteristic variables within the selected dataset, collected in the context of mental health issues among workers in an anonymous workplace, significantly influences the hyperparameter tuning, classification performance, and runtime of the ML model. In this context, it is important to determine and apply the right model and make the right intervention early to increase the performance of employees in the workplace.

Thereby our study not only addresses a critical gap in literature but also present a viable approach for the workplace directors by exploring the mental health status of individuals in the workplace through six mainstream ML algorithms.

2. Material and Method

The dataset used was structured in a tabular format, comprising 27 feature columns and 1,257 rows of individual data. It is crucial to emphasize at this point that the classification phase may present considerable challenges for the ML algorithms, and training the selected architecture with samples that lack adequate labeling information can lead to unreliable or meaningless predictions from the algorithm. The columns "Timestamp", "Country", "State" and "Comments" in the dataset were excluded from the analysis due to their minimal impact on the desired outcome. While country information could have provided valuable insights, the dataset's uneven distribution across countries limited its utility. The United States has the largest number of participants (751), followed by the United Kingdom (185) and Canada (72), while several other countries, such as Spain, Japan, and Nigeria, are represented by a single participant each. Drawing conclusions about a specific country's greater struggle with employee mental health would be misleading, as nearly 60% of the employees are citizens of the United States. Note that a brief review of the "States" column reveals its relevance only for respondents from the United States, prompting its exclusion from the analysis. Additionally, the "Comment" column contains a high proportion of missing values (70%), which is understandable given that it was an optional text field, and it is reasonable to expect that many respondents would leave it blank. Hence, a set of Python codes was employed for the data cleansing process, which involved standard procedures such as filling in missing data and addressing outliers. Subsequently, in the dataset under consideration, certain columns contained misspelled words (e.g., "femake" instead of "female" and "mail" instead of "male") and missing values (1% in the "self_employee" column and 21% in the "work_interfere" column). All of the misspellings in the data were corrected carefully, and the missing parts of the data were imputed with the most frequently occurring values. Subsequently, to identify an ML relationship between the independent variables and the selected attribute, categorical data must be transformed into numerical values using dummy variables. So, one can see that Table 1 provides variables and introductory information related to the individuals' conditions, including age, gender, employment status, family history of mental health, and prior treatment for mental health conditions. On the other hand, Table 2 presents parameters that may be associated with mental health status, derived from individuals' workplace environment information. Furthermore, in our ML analysis, the attribute "treatment" is used as the target feature, while the other features presented in Tables 1 and 2 are considered as predictors. The reader can also see the numerical counterparts of the categorical data in these tables to be used in the ML analyses.

Table 1. Demographic Variables extracted from the selected mental health data

Variable	Meaning	Values
Age	Age	Below 45 (0), 46-55 (1), Above 56 (2)
Gender	Gender	Male (0), Female (1), Non-Binary (2)
self_employed	Operating own business	No (0), Yes (1)
family_history	Family history of mental health	No (0), Yes (1)
Treatment	Received care for any mental health condition	No (0), Yes (1)

Table 2. Workplace factors extracted from the selected mental health data

Factor	Meaning	Values
work_interfere	Effect of mental state on performance	Sometimes (0), Often (1), Rarely (2), Never (3)
no_employees	Number of employees	1-5 (0), 6-25 (1), 26-100 (2), 101-500 (3), 501-100 (3), >1001 (4)
remote_work	Doing work remotely	No (0), Yes (1)
coworkers	Talking about a mental health concern with coworkers	No (0), Yes (1), Sometimes (2)
benefits	Benefits offered by employer	No (0), Yes (1), Don't know (2)
care_options	Care options available	No (0), Yes (1), Not Sure (2)
well_prog	Worker wellness program	No (0), Yes (1), Don't know (2)
seek_help	Support for mental challenges	No (0), Yes (1), Don't know (2)
anonymity	Anonymity	No (0), Yes (1), Don't know (2)
obs_conseq	Observed mental issues	No (0), Yes (1)
mh_conseq	Hiding mental health condition	No (0), Yes (1), Maybe (2)
ph_conseq	Hiding physical health condition	No (0), Yes (1), Maybe (2)
leave	Take time off for mental health reasons	Don't know (0), Very Easy (1), Somewhat Easy (2), Very Difficult (3), Somewhat Difficult (4)
tech_company	Tech-based organization	No (0), Yes (1)
supervisor	Talking with supervisors	No (0), Yes (1), Some of Them (2)
mh_interview	mental health concern in an interview with potential employer	No (0), Yes (1), Maybe (2)
ph_interview	Physical health concern in an interview with potential employer	No (0), Yes (1), Maybe (2)
ment_vs_phys	employer treats mental issues with the same level of importance as physical health case	No (0), Yes (1), Don't know (2)

According to traditional statistics, the mental health status of an employee shows a significant correlation with the features gender, family_history, work_interfere, coworkers, benefits, care_options, well_prog, seek_help, anonymity, obs_conseq, mh_conseq, leave, mh_interview, and ment_ve_phys. However, no significant differences were found in relation to Age (p-value is 0.449), self_employed (p-values is 0.625), no_employed (p-values is 0.137), remote_work (p-values is 0.384), ph_conseq (p-values is 0.246), tech_company (p-values is 0.283), supervisor (p-values is 0.420), and ph_interview (p-values is 0.203).

We have now reached the stage where we can provide information about the method we follow for the ML algorithms. The sample-based ML architectures are computationally simple and are frequently associated with various models of human learning [26]. In this context, Instance-Based Learning (IBL) is a fundamental ML method that uses the nearest neighbor approach [27]. In Figure 1, we present the flow chart of classifier modeling.

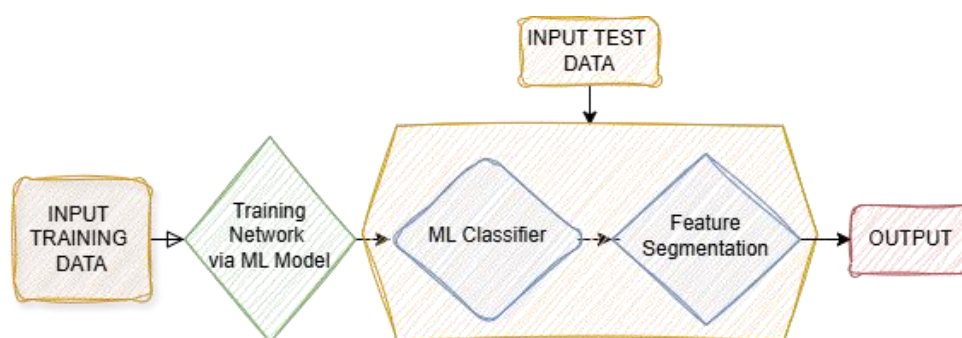


Figure 1. The workflow of classifier modeling

In the present investigation, we consider six different ML architectures in order to perform classification analysis within the scope of this study:

- **Bootstrap Aggregating (Bagging) Classifier:** Developed by Breiman in 1996 [28], this method is designed to improve the performance of models with high variance, such as decision trees.
- **Extremely Randomized Trees (ExtraTrees) Classifier:** This model, a variant of the decision tree approach, seeks to enhance prediction accuracy and generalizability by introducing a higher degree of randomness [29].
- **Passive-Aggressive Classifier:** This architecture, developed by Crammer, Dredze and Pereira in 2006 [30], is an online learning method that facilitates continuous updates to predictions by adapting to data streams, especially in large datasets.
- **Light Gradient Boosting Machine (LGBM) Classifier:** This architecture was developed by a team led by Ke [31] under the auspices of Microsoft Research. It was designed to address the limitations of traditional gradient boosting algorithms, offering an effective solution particularly for large-scale datasets and high-dimensional problems. By leveraging advanced techniques such as histogram-based feature binning and a leaf-wise tree growth strategy, the LGBM achieves superior computational efficiency and predictive performance, making it a powerful tool for modern machine learning challenges.
- **Perceptron Classifier:** The Perceptron model, developed by Rosenblatt in 1958 [32], is regarded as one of the fundamental building blocks of artificial neural networks and represents a significant milestone in the history of ML and AI. Inspired by the neural processes in the human brain, the model was constructed as a learning framework capable of performing classification tasks on linearly separable datasets.
- **Stochastic Gradient Descent (SGD) Classifier:** This is an efficient optimization method, particularly for large-scale datasets and high-dimensional problems [33]. By utilizing mini-

batches or individual data samples for weight updates, it significantly reduces memory usage and computational costs, making it highly suitable for scalable ML tasks.

The performance of ML processes can be assessed using a variety of evaluation metrics [34]. One of the most commonly utilized metrics is the confusion matrix, which is organized in a manner that enables the identification of true (T) and false (F) values derived from the classification results. Note that the confusion matrix offers a thorough assessment of a model's performance across various classes. In binary classification tasks, it is commonly represented as a 2×2 matrix, as shown in Figure 2.

		Predicted Values	
		Positive (P)	Negative (N)
Actual Values	Positive (P)	True Positive (TP)	False Negative (FN)
	Negative (N)	False Positive (FP)	True Negative (TN)

Figure 2. Main structure of a confusion matrix

Thence, the performance of an ML algorithm can be assessed through a variety of metrics, including accuracy, precision, recall, and the F1-score [35]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total Instances}}, \quad (2.1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (2.2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2.3)$$

$$\text{F1 - score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (2.4)$$

Accuracy, also referred to as the recognition rate, is taken into account as the ratio of correctly classified test instances to the total number of test samples, while the average accuracy across all classes reflects the classifier's overall performance. Precision quantifies the exactness of a classifier, representing the proportion of instances classified as positive that are truly positive. It reflects the classifier's capacity to minimize the incorrect labeling of negative samples as positive. Recall assesses the completeness of a classifier, indicating the proportion of actual positive instances that are correctly identified. This interesting metric reflects the classifier's ability to capture all positive samples. The F1-score is the harmonic mean of precision and recall, offering a balanced metric that accounts for the trade-off

between these two performance measures. This quantity can be useful in scenarios involving imbalanced datasets, as it incorporates both false positives and false negatives. The performance evaluation metrics discussed here are crucial for determining which of the machine learning approaches we compare in this study yields the most successful results on the mental health data under consideration.

3. Data Analysis

The chosen mental health dataset was partitioned into training (879 employees, 70%) and test (378 employees, 30%) parts. To evaluate the performance of diverse ML techniques, we examine individual features from the dataset, where each category represents the mental health status of an employee. This approach resulted in two distinct groups: one including individuals with mental health issues and the other comprising those without. This type of grouping is frequently used in comparative analyses to examine the role of each feature and its effect on the performance of the chosen ML approach. To further strengthen our investigation, we also examine the influence of features on the outcome variable. It is known that feature importance refers to the process of identifying which attributes within a dataset have the greatest influence on a model's estimations [36]. This process aids in pinpointing the most significant factors for predicting the target outcome, allowing the model to focus on these critical features. Understanding feature importance is crucial for improving model performance by facilitating more effective feature selection, enhancing interpretability, and mitigating the risk of overfitting through the removal of irrelevant or less impactful features. In this study, the Python environment, coupled with the Jupyter Notebook interface [37], was employed to analyze the data and assess feature importance. So, Table 3 outlines the chosen values for key hyperparameters in the ML classifiers utilized in this study. These selections are based on commonly reported values in the existing literature and were refined through several preliminary tests.

Table 3. Hyperparameters of the selected ML architectures

Classifier	Hyperparameter	Value
Bagging and ExtraTrees	Number of Random State	42
	Number of Estimators	100
Passive-Aggressive, Perceptron and SGD	Maximum Iteration	100
	Number of Random State	42
	Stopping Criterion	0.001
LGBM	Learning Rate	0.09
	Number of Classes	2
	Number of Random State	42

Figures 3 and 4 present the performance results of the chosen ML architectures applied to our dataset. As previously noted, accuracy, precision, recall, and the F1-score are the metrics used to evaluate the performance of an ML model. Focusing on the accuracy values, it is clear that the Bootstrap Aggregating (Bagging) classifier surpasses the other ML approaches in performance on the given dataset, attaining a value of 0.74. The performance of the other models, in descending order, is as follows: the Light Gradient Boosting Machine (LGBM), Extremely Randomized Trees (ExtraTrees), Perceptron, Stochastic Gradient Descent (SGD) and the Passive-Aggressive. Given that accuracy is a fundamental metric for evaluating classification performance, the results presented in Figure 4 indicate that the Bootstrap Aggregating (Bagging) regressor yields the most effective model fit for the selected mental health dataset. Likewise, when considering the additional metrics of precision, recall, and F1-score, the results presented in Figure 4 affirm that the Bagging Classifier model emerges as the highest-performing approach for our dataset.

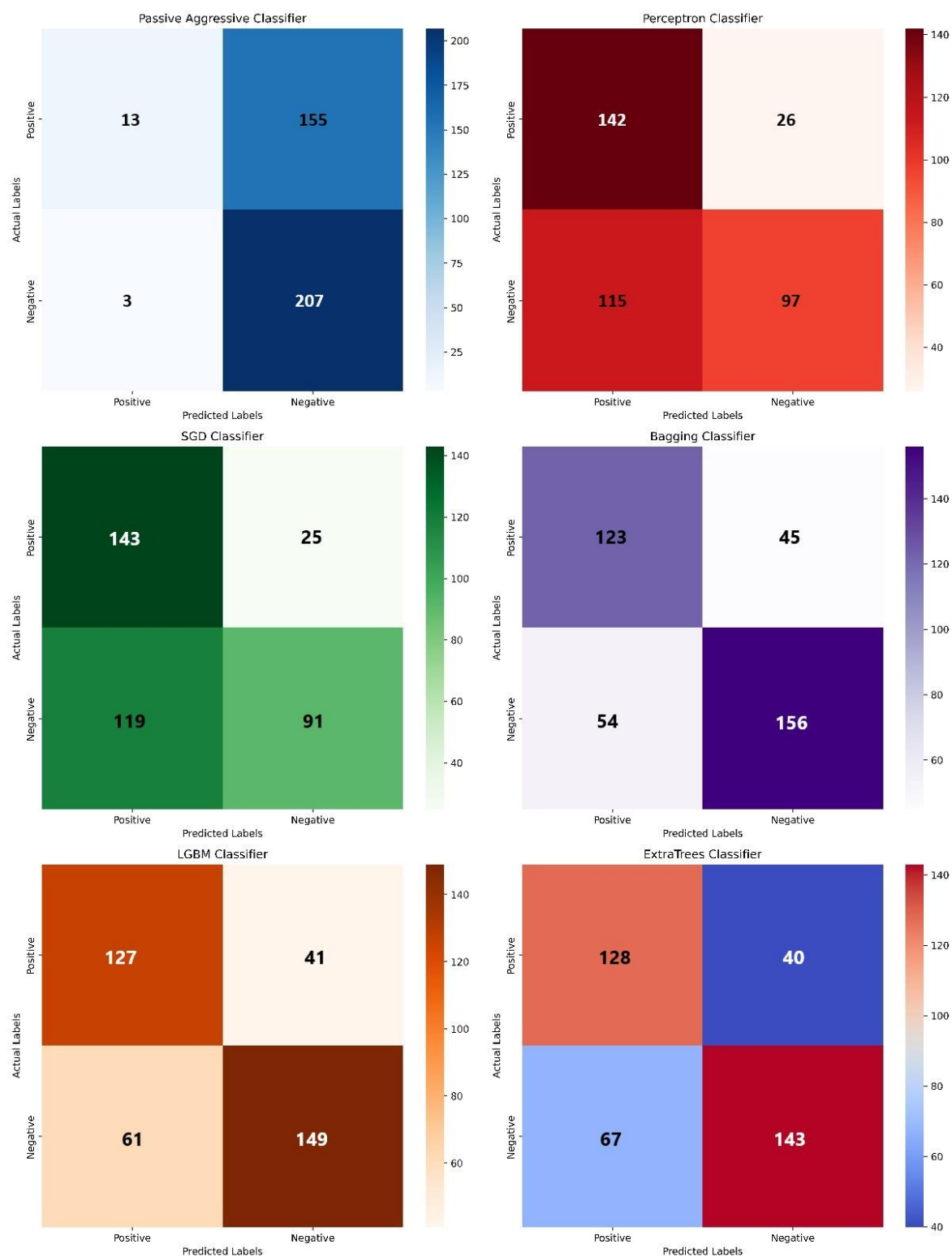


Figure 3. Confusion matrices for the ML models

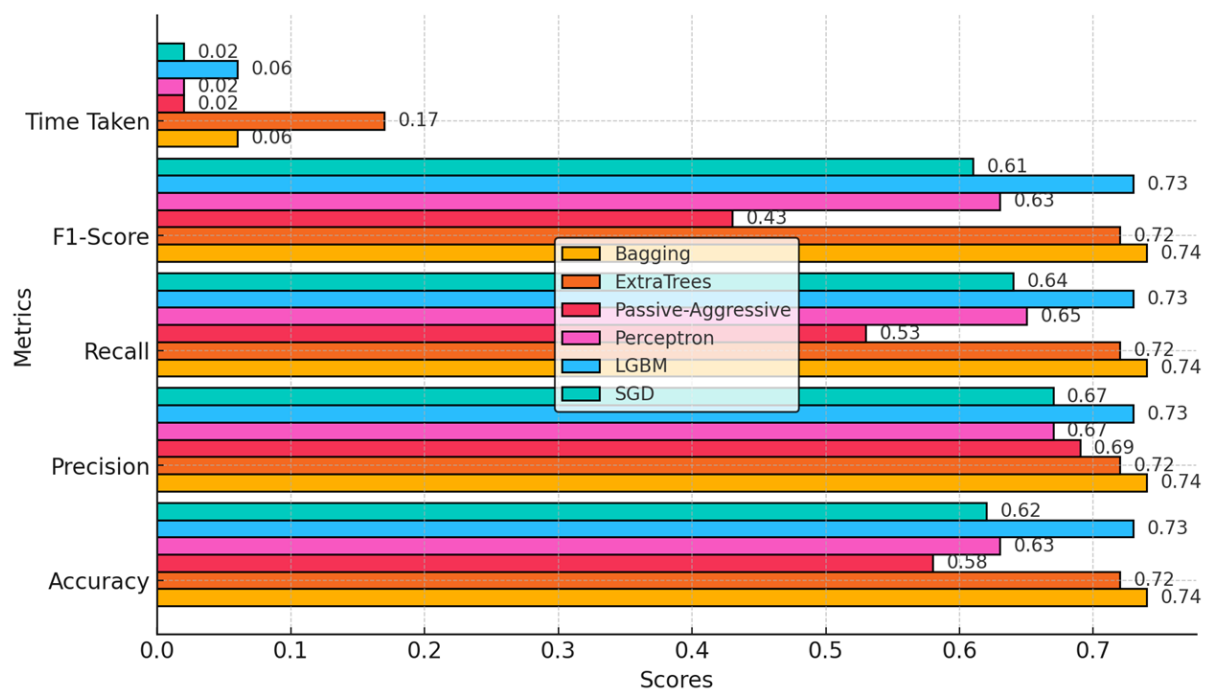


Figure 4. Performance results for the selected ML classifiers

Next, Figure 5 presents the Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves, which illustrate the comparative performance of the ML models employed for the binary classification task under investigation. It is evident that the Bagging Classifier achieves the highest Area Under the Curve (AUC) score of 0.79, alongside an Average Precision (AP) score of 0.79. These metrics demonstrate that the Bagging model outperforms the others in effectively discriminating between the positive and negative classes. In contrast, the Passive-Aggressive model exhibits the lowest performance among the evaluated models, with an Area Under the Curve (AUC) of 0.72 and an Average Precision (AP) of 0.72. These results indicate that the Passive-Aggressive classifier performs relatively poorly in distinguishing between the positive and negative classes compared to the other approaches.

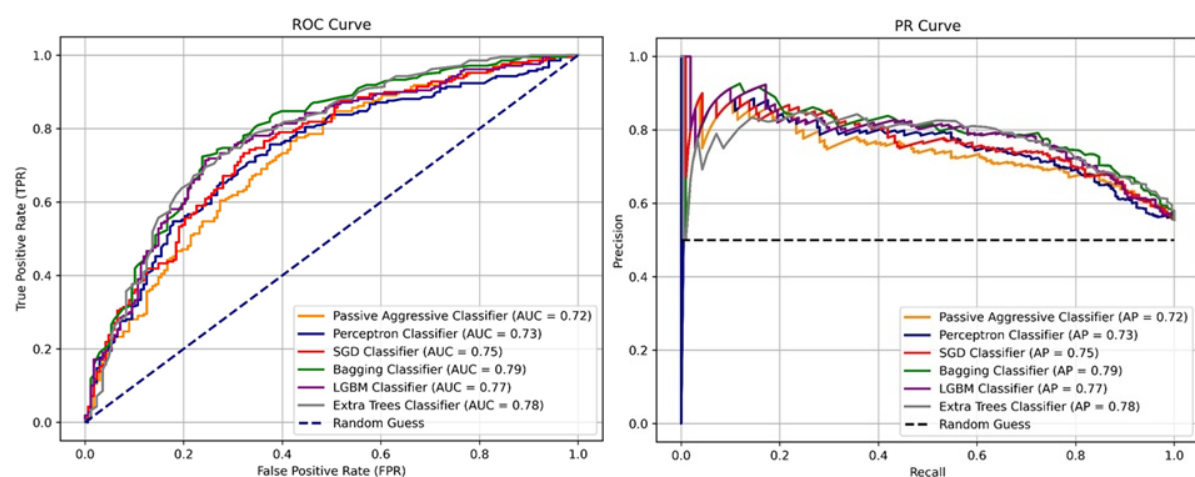


Figure 5. ROC (left) and PR (right) curves

In the subsequent phase of our research, we conduct a detailed examination of the specific contributions of each feature within the dataset. In this context, Figure 6 presents the results of feature importance analysis derived from the selected ML models. These analyses identify and rank the factors influencing

an individual's mental health. Based on the results obtained from the Bagging model, which demonstrates the highest accuracy rate, it is evident that the five most critical factors influencing an employee's mental state are work_interfere, family history, Age, no_employees, and care_options. Among individuals with a family history of mental health issues, the proportion of those receiving treatment (Yes) is notably higher than those not receiving treatment (No). This indicates that such a family history may play a crucial role in raising awareness and motivating individuals to seek professional help.

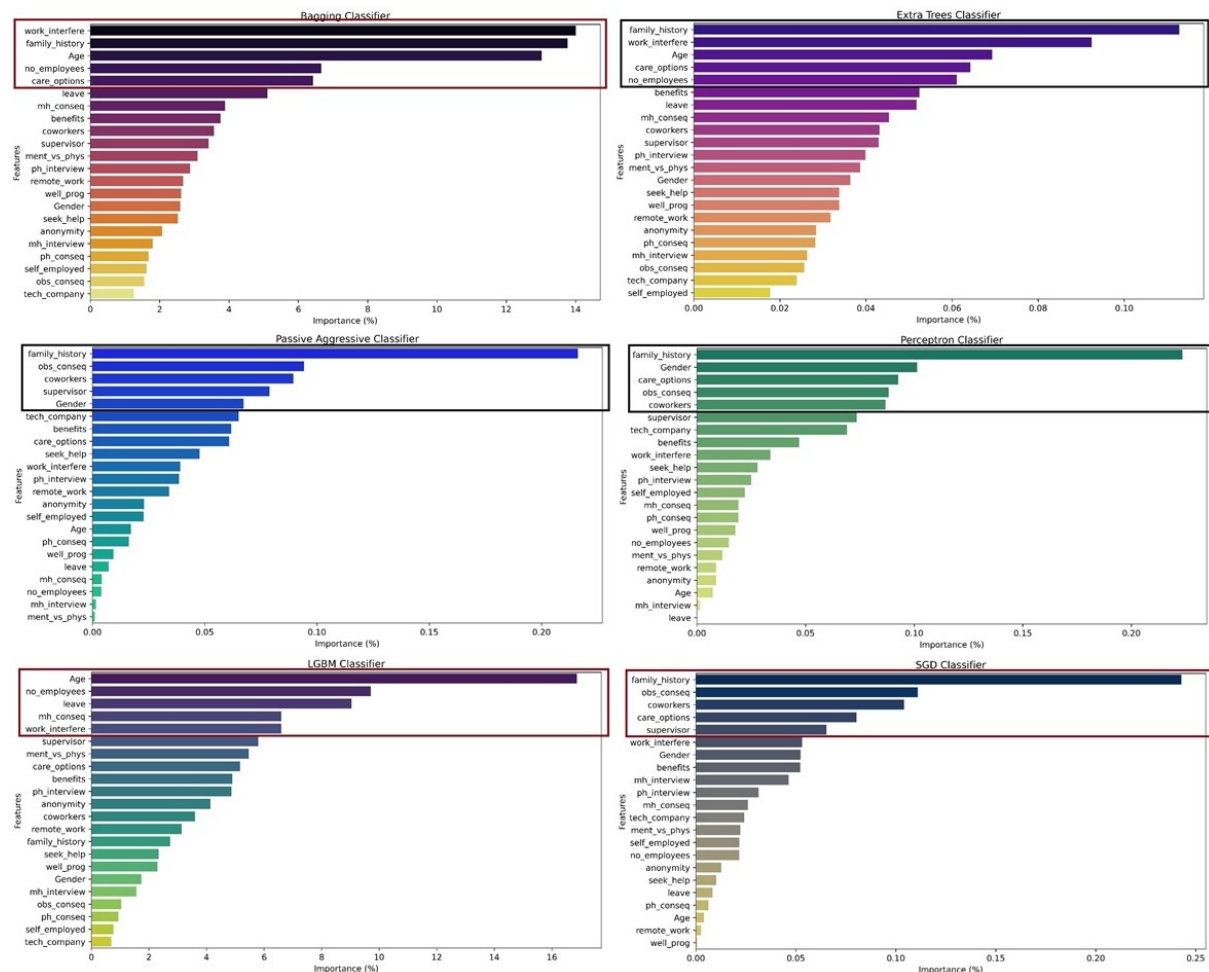


Figure 6. Illustration of our feature importance analysis

Moreover, prioritizing the factors that have the most significant impact on an individual's mental health, as identified by all models, will enable us to derive additional novel inferences. Considering all models, the factors coded as work_interfere, family history, age, no_employees, and care_options, along with obs_conseq, gender, coworkers, and leave, can be regarded as influential determinants of an individual's mental well-being level.

4. Clinical Prospects

The foundational step in conducting interdisciplinary research within behavioral science is to clearly establish the intended research objective. This typically involves selecting between two key computational strategies: forecasting a specific label or constructing a rule-based framework to define the label. Through the objective quantification of behavior using the ML algorithms, we can systematically track behavioral dynamics over extended periods. This creates significant opportunities, enabling the assessment of trajectories both preceding and following a diagnosis, as well as the evaluation of responses to

pharmacological or behavioral interventions. Moreover, as a component of a strategy to enhance a patient's behavioral health, offering healthcare professionals objective feedback on their practices through ML applications can substantially aid in their professional growth. Therefore, clinicians can become better equipped to track patient progress and adjust their approaches as needed through a detailed ML-based analysis of therapy and diagnostic sessions. These circumstances make the ML-based data analysis crucial within the context of clinical applications. In addition, an AI-driven diagnostic tool has the potential to identify diseases at their early stages by rapidly analyzing health data, thereby enabling healthcare professionals, including nurses and doctors, to expedite the treatment process and minimize errors. Consequently, the AI enhances decision-making processes, elevates the quality of care, and strengthens patient safety, all while alleviating the workload within healthcare settings. However, it is crucial to address considerations related to the ethical implementation of these technologies, safeguarding data privacy, and ensuring proper integration into clinical workflows.

Although we have highlighted the potential of ML algorithms in conducting behavioral analysis of employees to improve mental health, as the field progresses, key questions that require careful consideration emerge. For instance, it is vital that healthcare professionals receive comprehensive training to ensure the effective use of AI tools. In this regard, the ML-based researches serve as a valuable guide for the healthcare sector. Currently, a range of expert systems are being utilized in medicine to detect mental health disorders early, thereby enabling more precise and efficient treatment approaches. The insights derived from the successful analysis of mental health data hold significant value for healthcare professionals, who play a key role in offering guidance on strategies to manage stress, anxiety, and depression. With a variety of ML techniques available for constructing expert systems, it becomes imperative to assess and choose the most suitable method to address the particular health challenge. Within this framework, incorporating existing technologies into every facet of mental health research and care can bridge gaps in scientific knowledge while fostering a trusted environment to advance new intervention methods.

5. Discussions

In this research, we have examined the Bagging, ExtraTrees, Passive-Aggressive, Perceptron, LGBM and the SGD classifiers to effectively assess the risk factors related to mental health among employees. The results unequivocally demonstrate that the Bagging algorithm surpasses the other methods in terms of the accuracy, AUC and the AP scores. Following the analysis of the selected dataset using the aforementioned regressors, we have identified key predictors, including work interference, family history, age, number of employees and care options, thereby confirming their associations with mental health-related behaviors. The application of ML techniques has further refined these insights, allowing for a more thorough understanding of the data and a ranking of the predictive significance of the variables previously discussed.

There is no doubt that computational science has a significant role to play in the study of human behavior, making this an exciting period for the ML researchers. Clearly, addressing the major challenges in mental health research and clinical practice requires a collaborative effort. Moreover, key computational objectives encompass interaction modeling along with behavior prediction. Subsequently, enduring scientific advances in mental health domain can be achieved if we can overcome the technical obstacles. Beyond the practical advantages of data integration, the ML has the potential to provide valuable theoretical advancements in areas such as the theory of mind.

Acknowledgment

The authors would like to thank the reviewers and editorial boards of the *International Journal of Pure and Applied Sciences*.

Conflict of Interest

The authors declare no competing interests.

Research and Publication Ethics Statement

The data is sourced from an open-access database, so there is no need for an ethics committee's evaluation.

Data Availability Statement

The data used to support the findings of this study and the code used are available from the corresponding author upon reasonable request.

References

- [1] Doupe, P., Faghmous, J. and Basu, S. (2006). *Data mining: Introductory and advanced topics*, Pearson Education India.
- [2] Ghahramani, Z. (2004). Unsupervised learning. *Advanced lectures on machine learning*, Springer; pp. 72-112.
- [3] Dunham, M.H. (2006). *Data mining: Introductory and advanced topics*. Pearson Education India.
- [4] Kotsiantis, S.B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31, 249–268
- [5] Cortes, C. and Vapnik, V.N. (1995). Support Vector Network, *Machine learning*, 20, 1-25.
- [6] Panch, T., Szolovits, P. and Atun, R. (2018). Artificial intelligence, machine learning and health systems. *J. Glob. Health*, 8(2), 020303.
- [7] Doupe, P., Faghmous, J. and Basu, S. (2019). Machine Learning for Health Services Researchers. *Value in Health*, 22, 808-815.
- [8] Husnain, A., et al. (2024). Advancements in Health through Artificial Intelligence and Machine Learning: A Focus on Brain Health. *Revista Española de Documentación Científica*, 18(01), 100-123.
- [9] Gogas, P. and Papadimitriou, T. (2021). Machine Learning in Economics and Finance. *Computational Economics*, 57, 1-4.
- [10] Shami, L. and Lazebnik, T. (2024). Implementing Machine Learning Methods in Estimating the Size of the Non-observed Economy. *Computational Economics*, 63, 1459–1476.
- [11] Ntampaka, M., et al. (2019). A Deep Learning Approach to Galaxy Cluster X-Ray Masses. *ApJ*, 876, 82.
- [12] Kangal, E.E., Salti, M. and Aydogdu, O. (2019). Machine learning algorithm in a caloric view point of cosmology. *Physics of the Dark Universe*, 26, 100369.
- [13] Tilaver, H., Salti, M., Aydogdu, O. and Kangal, E.E. (2021). Deep learning approach to Hubble parameter. *Computer Physics Communications*, 261, 107809.
- [14] Salti, M., Kangal, E.E. and Aydogdu, O. (2021). Evolution of CMB temperature in a Chaplygin gas model from deep learning perspective. *Astronomy and Computing*, 37, 100504.
- [15] Chase, R.J., et al. (2022). A Machine Learning Tutorial for Operational Meteorology - Part I: Traditional Machine Learning. *Weather and Forecasting*, 37(8), 1509–1529.
- [16] Buster, G., Benton, B.N., Glaws, A. and King, R.N. (2024) High-resolution meteorology with climate change impacts from global climate model data using generative machine learning. *Nature Energy*, 9, 894-906.
- [17] Sumathi, M.R. and Poorna, B. (2016). Prediction of Mental Health Problems Among Children Using Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications*, 7(1), 552-557.
- [18] Powers, S.I., Hauser, S. T. and Kilner, L.A. (1989). Adolescent mental health. *American Psychologist*, 44(2), 200–208.
- [19] Bhugra D., Till A. and Sartorius N. (2013). What is mental health? *International Journal of Social Psychiatry*, 59(1), 3-4.

- [20] Reddy, U.S., Thota, A.V. and Dharun, A. (2018). Machine Learning Techniques for Stress Prediction in Working Employees, *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, (pp. 1-4), Madurai, India.
- [21] Katarya, R. and Maan, S. (2020). Predicting Mental health disorders using Machine Learning for employees in technical and non-technical companies, *2020 IEEE International Conference on Advances and Developments in Electrical and Electronics Engineering (ICADEE)*, (pp. 1-5), Coimbatore, India.
- [22] Jain, T., Jain, A., Hada, P.S., Kumar, H., Verma, V. K. and Patni, A. (2021). Machine Learning Techniques for Prediction of Mental Health, *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, (pp1 1606-1613), Coimbatore, India.
- [23] Garriga, R., Mas, J., Abraha, S. et al. (2022). Machine learning model to predict mental health crises from electronic health records. *Nat Med* 28, 1240–1248.
- [24] Rezapour, M. and Hansen, L. (2022). A machine learning analysis of COVID-19 mental health data. *Scientific Reports*, 12, 14965.
- [25] Mulye, A. (2021). Mental Health at Workplace. Kaggle.
- [26] Hall, M.A. (1999). Correlation-based Feature Selection for Machine Learning. PhD Thesis / University of Waikato, Hamilton, New Zealand.
- [27] Cunningham, S.J., Littin, J. and Witten, I.H. (1997). Applications of machine learning in information retrieval. University of Waikato Technical Report 97/6.
- [28] Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- [29] Geurts, P., Ernst, D. and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42.
- [30] Crammer, K., Dredze, M., & Pereira, F. (2006). Passive-Aggressive Algorithms for Huge Online Tasks. *Journal of Machine Learning Research*, 7, 551-585.
- [31] Ke, G. et al. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA.
- [32] Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6), 386-408.
- [33] Bottou, L. (1998). Stochastic Gradient Descent. *Online Learning and Neural Networks*, MIT Press.
- [34] Caruana, R., Niculescu-Mizil, A. (2004). Data mining in metric space: an empirical analysis of supervised learning performance criteria. *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, Seattle-WA, USA.
- [35] Labatut, V. and Cherifi, H. (2011). Evaluation of Performance Measures for Classifiers Comparison. *Ubiquitous Computing and Communication Journal*, 6:21-34.
- [36] Chen, T., et al. (2019). Prediction of Extubation Failure for Intensive Care Unit Patients Using Light Gradient Boosting Machine. *IEEE Access*, 7, 150960-150968.
- [37] Project Jupyter. Jupyter Notebook: The Classic Notebook Interface. Available at the website <https://jupyter.org>.