Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 9 No. 2 (2025) 185 - 194 e-ISSN: 2580-0760

Comparative Analysis of Machine Learning Algorithms for Predicting Patient Admission in Emergency Departments Using EHR Data

Ahmad Abdul Chamid[1*], Ratih Nindyasari[2], Muhammad Imam Ghozali[3]
[1,2,3]Department of Informatics Engineering, Faculty of Engineering, Universitas Muria Kudus, Kudus, Indonesia
[1]abdul.chamid@umk.ac.id, [2]ratih.nindyasari@umk.ac.id, [3]imam.ghozali@umk.ac.id

Abstract

Every patient who is rushed to the Emergency Department needs fast treatment to determine whether the patient should be inpatient or outpatient. However, the existing fact is that deciding whether an inpatient or outpatient must wait for the diagnosis made by the existing doctor, so if there are many patients, it generally takes quite a long time. So, to predict patient admissions to the emergency unit, a machine learning model that can be fast and accurate is needed. Therefore, this study developed a machine learning and neural network model to determine patient care in Emergency Departments. This study uses publicly available electronic health record (EHR) data, which is 3,309. The model development process uses machine learning methods (SVM, Decision Tree, KNN, AdaBoost, MLPClassifier) and neural networks. The model that has been obtained is then evaluated for its performance using a confusion matrix and several matrices such as accuracy, precision, recall, and F1-Score. The results of the model performance evaluation were compared, and the best model was obtained, namely the MLPClassifier model with an accuracy value = 0.736 and an F1-Score value = 0.635, and the Neural Network model obtained an accuracy value = 0.724 and an F1-Score value = 0.640. The best models obtained in this study, namely the MLPClassifier and Neural Network models, were proven to be able to outperform other models.

Keywords: *Emergency Departments; Electronic Health Record; Machine Learning; Neural Networks; Patient Care*

How to Cite: A. A. Chamid, R. Nindyasari, and M. I. Ghozali, "Comparative Analysis of Machine Learning Algorithms for Predicting Patient Admission in Emergency Departments Using EHR Data", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 9, no. 2, pp. 185 - 194, Mar. 2025.
DOI: https://doi.org/10.29207/resti.v9i2.6188

1. Introduction

The large number of patients entering the Emergency Department (ED) will result in a crowded room and become a challenge in its management. The problems that arise are limited space, lack of staff, and potential loss of revenue. Limited space when there are many patients in the ED hinders patient flow and increases waiting time. So, it will create a crowded waiting room and increase frustration for patients and staff. Generally, to overcome these obstacles, it is often necessary to utilize non-traditional areas such as hallways to accommodate patients. This emergency solution will hurt patient satisfaction, lower patient experience scores, and create a sense of disorganization for visitors [1], [2].

Every person rushed to the hospital needs a quick diagnosis to determine whether the patient needs to be an inpatient or outpatient. Generally, to decide whether a patient is an inpatient or outpatient, the doctor on duty must wait for the decision based on observations made on patients who enter the Emergency Department (ED). The medical field can utilize technology development by electronically recording health records for every patient who comes to the ED. The results of each patient's health record can be processed quickly to determine whether the patient needs to be an inpatient or outpatient using artificial intelligence. Machine learning is a part of artificial intelligence that can be used to process data as model training [3]; it can also be used to build models by utilizing a large amount of available data [4].

The development of machine learning has been applied in the medical field, including determining the estimated duration of home health services; Decision Tree algorithms, Random Forest, and Multi-layer Perceptron Neural Network algorithms are used to build estimation models [5]. In reality, the health service budget is limited, so it requires optimal resources;

Received: 25-11-2025 | Accepted: 26-02-2025 | Published Online: 07-03-2025

machine learning has great potential to utilize health service resources by being used to predict health service costs in acutely ill patients [6]. Machine learning can tackle healthcare information overload [5]. It is proven that artificial intelligence (AI) and machine learning (ML) contribute to a wealth of information. Still, the drawbacks are that training data needs to be peer-reviewed, raises ethical and legal issues, and needs more human empathy. On the other hand, there is optimism that AI, including ML, will help manage complex information by processing data, determining diagnoses, recommending treatments, and predicting outcomes.

Research related to the use of machine learning for health continues to grow, including ML being used to predict type 1 diabetes in children [8] In general, the views of doctors and the public towards the use of machine learning-based prediction models are positive. Globally, the diagnosis of childhood type 1 diabetes is still often delayed despite significant campaigns to raise awareness among the public and primary health care practitioners. Research suggests that children can be identified earlier using machine learning algorithms. To assess the effectiveness of such algorithms in practice, the feasibility and acceptability of the tool must be evaluated in primary healthcare settings [9]. In general, risk prediction models using ML are viewed positively. Healthcare professionals and patients believe that these models have the potential to provide additional benefits in the healthcare environment. However, concerns remain, such as issues with the data quality used to develop the models and fears of possible unintended consequences after implementing ML models [10].

AI applications utilizing machine learning (ML) models in healthcare have increased [11]. The ML model developed requires an algorithm appropriate to the case study being worked on. Previously, there were several algorithms used to create ML models in the health sector, including neural networks [12], [13] A neural network (NN) is used to develop an artificial intelligence model in detecting peripheral artery disease (PAD) based on electronic health records (HER), the performance of NN combined with EHR data shows the feasibility of identifying PAD cases before diagnosis [14]. The Random Forest (RF) algorithm is used to predict mortality risk in pediatric intensive care units (PICU) in hospitals [15], and the results show a precision-recall value of 0.78. In addition, RF is used to predict various case studies in health services [16], [17]. Furthermore, the Naive Bayes (NB) algorithm is used for suicide risk prediction in hospital clinical practice [18]. In addition, the XGBoost algorithm is used for breast disease recurrence prediction [19].

Machine learning algorithms have been widely applied in healthcare services, namely the use of K-Nearest Neighbor (KNN) to classify fetal health using cardiotocogram data [20], parameter numbers of neighbors with a range of 1-31, applying three experimental scenarios resulted in the best accuracy value of 0.97 in the last scenario. In addition, the Artificial Neural Network (ANN), XG Boost (XGB), Support Vector Machine (SVM), Light GBM (LGBM), and Random Forest (RF) algorithms have been used to develop machine learning models [21], [22], [23], [24], [25], [26]. Apart from that, it has been tested. Its performance is compared to get an accuracy value of 0.89-0.99 [20]. Next, machine learning is used to automatically classify CTG with a fetal medical record dataset from UCI, consisting of 2,126 data and 21 features. The category labels are Normal (N), Suspected (S), and Pathological (P). Methods such as Random Forest, SVM, Decision Tree, and K-Nearest Neighbour are used, with Accuracy, F1-Score, and ROC reaching more than 90% [27].

Based on previous research, it can be identified that the public's view of AI models may be more damaging than that of HCPs and that concerns (e.g., additional workload demands) are only sometimes borne out in practice. This conclusion is influenced by the low number of studies with patients and the public, the lack of ethnic diversity of participants, and the variation in the quality of articles. Furthermore, gaps in knowledge (particularly the views of underrepresented groups) and optimal methods for model explanation and warning require future research [10]. Therefore, this study developed a model to determine treatment for patients entering Emergency Departments based on health record data using Machine Learning and Neural Networks. The novelty of this study will test various machine learning and neural network algorithms, adjust parameters during the model-building process, and then compare the results of the model performance evaluation to obtain the best model.

2. Research Methods

This study uses publicly available patient health record data, and the original data can be accessed via the following link: https://data.mendeley.com/datasets/7kv3rctx7m/1 [28]. The data obtained is used to develop a classification model for determining patient care. The model development process uses the Python programming language. The research stages are shown in Figure 1.

The research stages, as shown in Figure 1, can be explained as follows:

This study uses public data like the Electronic Health Record (EHR) dataset. The original EHR data has been corrected and is publicly available at the following link: https://www.kaggle.com/datasets/manishkc06/patient-treatment-classification/data. 3.309 data were obtained, examples of which can be seen in Table 1.

This is an open access article under the CC BY-4.0 license

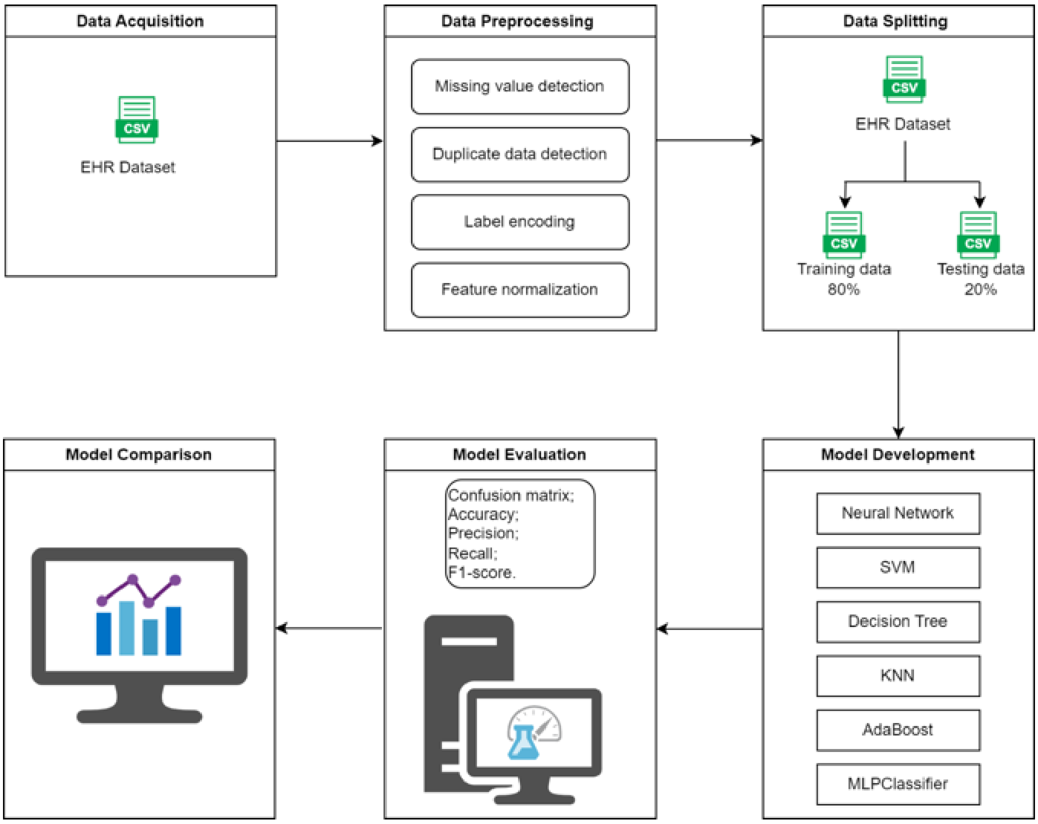

Figure 1. Research Stages

Table 1. The Example of Dataset

No	Haematocrit	Haemoglobins	Erythrocyte	Leucocyte	Thrombocyte	MCH	MCHC	MCV	Age	Sex	Source
0	33.8	11.1	4.18	4.6	150	26.6	32.8	80.9	33	F	1
1	44.6	14.0	6.86	6.3	232	20.4	31.4	65.0	36	M	0
2	42.9	14.0	4.57	6.2	336	30.6	32.6	93.9	70	F	0
3	41.9	14.4	4.67	3.5	276	30.8	34.4	89.7	18	F	0
4	40.6	13.3	4.85	14.9	711	27.4	32.8	83.7	36	M	0

Attributes in the dataset consist of attribute names, attribute types, measurement units, and brief descriptions. The following is the attribute information in the dataset (Name / Data Type / Value Sample / Description):

HAEMATOCRIT /Continuous /35.1 / Patient laboratory test result of hematocrit

HAEMOGLOBINS/Continuous/11.8 / Patient laboratory test result of hemoglobin’s

ERYTHROCYTE/Continuous/4.65 / Patient laboratory test result of erythrocyte

LEUCOCYTE /Continuous /6.3 / Patient laboratory test result of leucocyte

THROMBOCYTE/Continuous/310/ Patient laboratory test result of thrombocyte

MCH/Continuous /25.4/ Patient laboratory test result of MCH

MCHC/Continuous/33.6/ Patient laboratory test result of MCHC

MCV/Continuous /75.5/ Patient laboratory test result of MCV

AGE/Continuous/12/ Patient age

SEX/Nominal – Binary/F/ Patient gender

SOURCE/Nominal/ {1,0}/The class target 1 = in care patient, 0 = out care patient

Based on Figure 2, the distribution of continuous variables in the data set can be seen, and there are a thousand patients in the highest range of each variable. Patients have an average haemoglobin value = 14. In addition, the distribution of patient age is seen in the range of 1 to less than 100 years. Meanwhile, Figure 3 shows that the gender distribution is 1.566 female patients and 1.743 male patients. In addition, the class distribution is seen as 1.992 labeled as out care (0) and 1.317 labeled as in care.

This is an open access article under the CC BY-4.0 license

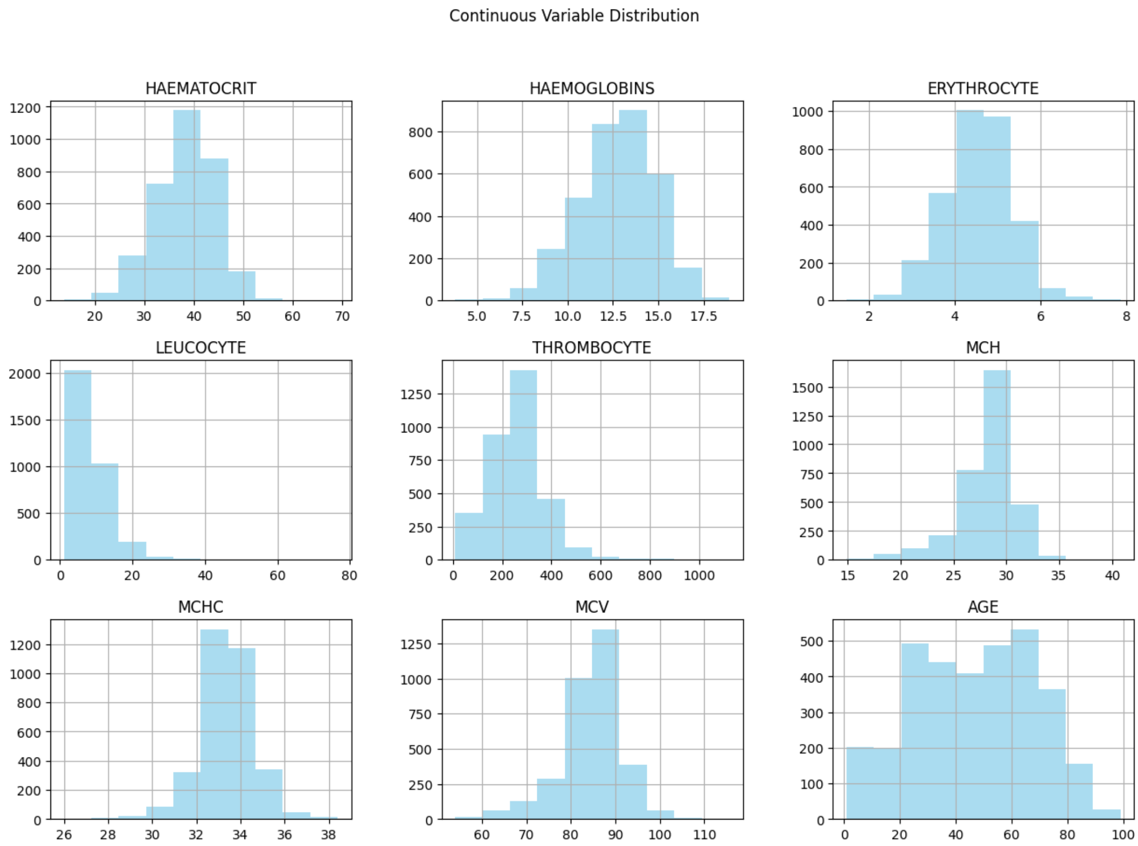

Figure 2. The Continuous Variable Distribution

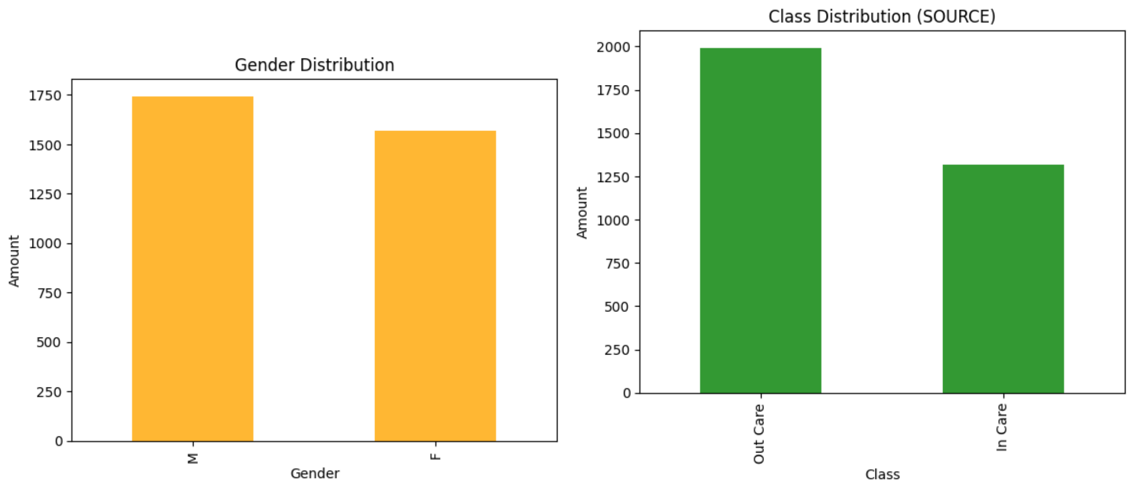

Figure 3. The Gender and Class Distribution

The data obtained is then pre-processed, and the stages carried out begin by checking the missing value and whether the available data contains empty values in certain attributes; after checking, it turns out that there are no missing values. Furthermore, duplicate data detection is carried out to check whether there is data duplication; after checking, it turns out that there is no data duplication. The next preprocessing step is to perform label encoding, which is a technique for converting categorical values into numeric form, which is required by machine learning algorithms that can only work with numeric data. On the SEX attribute, label encoding is performed (M = 1, F = 0). The last preprocessing step performs feature normalization, which is the process of re-scaling feature values so that they are in the same range, usually between 0 and 1 or -1 and 1. The goal is to ensure that all features have a uniform scale so that no feature dominates another simply because it has a larger value range.

At this stage, the data is divided with a percentage of 80% as training data and 20% as testing data.

The stage of developing the model using 6 algorithms: Neural Network (NN), Support Vector Machine

This is an open access article under the CC BY-4.0 license

(SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), AdaBoost, and MLPClassifier [29], [30], [31], [32]. The model development process is done by training data on each algorithm so that 6 models are obtained.

Each developed model is evaluated for its performance using the Confusion matrix, Accuracy, Precision, Recall, and F1-score [33].

After obtaining the evaluation results for each model, the next step is to compare its performance. The accuracy and F1-score matrices are used as references to determine the best model based on the highest value.

3. Results and Discussions

The results of this research are in the form of a machine learning model that has been built using several algorithms, including:

3.1 Neural Network (NN) Model

The neural network model was evaluated using the confusion matrix, the results of which are shown in Figure 4.

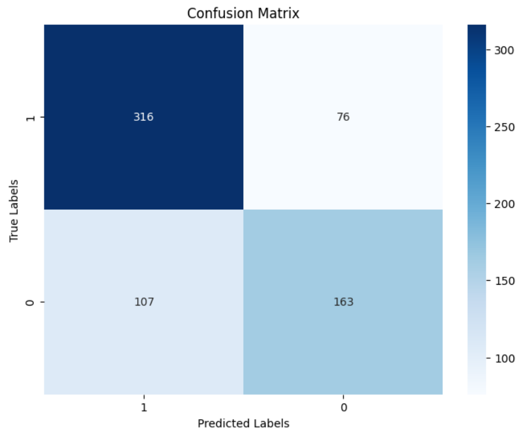

Figure 4. The Results of the Confusion Matrix of the Neural Network Model

The NN model confusion matrix results show that 107 data labelled as outpatient (0) are predicted as inpatient (1). There are 316 data labelled as an inpatient (1) that can be predicted as an inpatient (1). Furthermore, there are 76 data labelled as inpatient but predicted as outpatient, and there are 163 data labelled as outpatients successfully predicted as an outpatient. The NN model is evaluated using the accuracy, precision, recall, and F1-score matrices; the results are shown in Table 2.

Table 2. The Results of Neural Network Model Evaluation

Accuracy	Precision	Recall	F1-Score
0.724	0.682	0.604	0.640

Based on the results of the NN model evaluation shown in Table 2, it is known that the accuracy value = 0.724, then the F1-score value is known = 0.640. The accuracy value looks higher than the F1-score value; successively, the evaluation process is carried out from the accuracy value to the F1-score, which is the result of the calculation of the precision and recall values [34], [35].

3.2 Support Vector Machine (SVM) Model

The SVM model is built using training data and then evaluated using a confusion matrix, the results of which are shown in Figure 5.

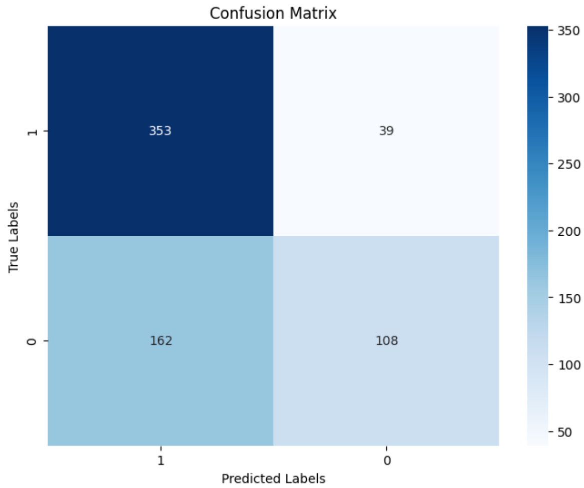

Figure 5. The Results of the Confusion Matrix of the SVM Model

The SVM model performance evaluation results using the confusion matrix, as shown in Figure 5, show that the model can predict correctly according to the data labeled as an inpatient (1) with as much as 353 data. Then, according to the data, the model can predict an outpatient (0) with as much as 108 data. While there are still errors in predicting 162 data that should be outpatient (0) predicted as inpatient (1). In addition, there are still errors in predicting 39 data that should be inpatient (1) and predicted as outpatient (0).

Furthermore, the SVM model's performance is evaluated using several evaluation matrices, the results of which are shown in Table 3.

Table 3. The Results of the SVM Model Evaluation

Accuracy	Precision	Recall	F1-Score
0.696	0.735	0.400	0.518

Based on the results of the SVM model performance evaluation, as shown in Table 3, it can be seen that the model obtained performance results with an accuracy value = 0.696, precision value = 0.735, recall value = 0.400, and F1-score value = 0.518. The evaluation results show that the precision value is higher than the others, but in this study, the reference used to see the model performance is the accuracy value and F1-score according to several previous sources [29], [33].

3.3 Decision Tree (DT) Model

After conducting data training using the DT algorithm, a model was obtained whose performance was then evaluated using a confusion matrix, the results of which are shown in Figure 6.

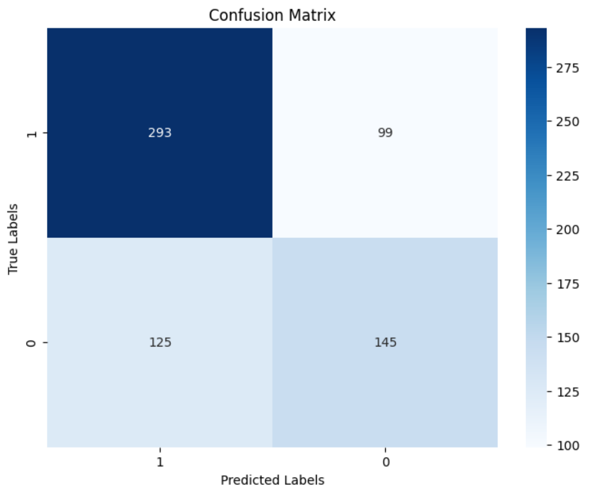

Figure 6. The Results of the Confusion Matrix of the DT Model

Based on the results of the DT model performance evaluation using the confusion matrix, as shown in Figure 6, it is known that the model can detect 293 data detected as inpatients (1) well. Then, 145 data, as well as outpatients (0), were detected. The model still has errors, detecting 125 data as inpatients (1), which should be detected as outpatients (0). In addition, 99 data points should be detected as inpatients (1), but the results are detected as outpatients (0). These results show how far the model can detect new data used for model testing.

Furthermore, several evaluation matrices are used to obtain the DT model evaluation results in more detail, which is shown in Table 4.

Table 4. The Results of DT Model Evaluation

Accuracy	Precision	Recall	F1-Score
0.662	0.662	0.537	0.564

The results of the DT model performance evaluation, as shown in Table 4, show that the accuracy value and precision value = 0.662, then the model gets the performance of the recall value = 0.537 and the F1-score value = 0.564. Based on the results of the DT model performance evaluation, it is known that the accuracy value is the highest value compared to the values of the other matrices. The results of the DT model performance seen in each matrix indicate that the model performance still needs to be improved, which can be done by adding training data [30], [31], [35].

3.4 K-Nearest Neighbor (KNN) Model

KNN is one of the classification algorithms commonly used to develop machine learning models. This study uses KNN to develop models by training data using the KNN algorithm. After training data, the next step is to obtain a KNN model that has been evaluated for its performance using a confusion matrix so that the results are as shown in Figure 7.

The results of the KNN model performance evaluation using the confusion matrix, as seen in Figure 7, show that the KNN model is able to predict well 331 data that are correctly predicted as inpatients (1). In addition, it can correctly predict 147 data that are correctly predicted as outpatients (0). Meanwhile, the weakness of the KNN model is that it has not been able to predict 123 data that should be outpatients (0) predicted as inpatients (1). In addition, the next weakness is that there is an error in predicting 61 data that should be inpatients (1) predicted as outpatients. Based on the evaluation using the confusion matrix, it can be seen that there are still several weaknesses in the KNN model, so it still needs to be improved. In addition to using the confusion matrix, several other matrices are needed for the model evaluation process to obtain more comprehensive model performance evaluation results. After evaluating the KNN model performance using several evaluation matrices, the results can be seen in Table 5.

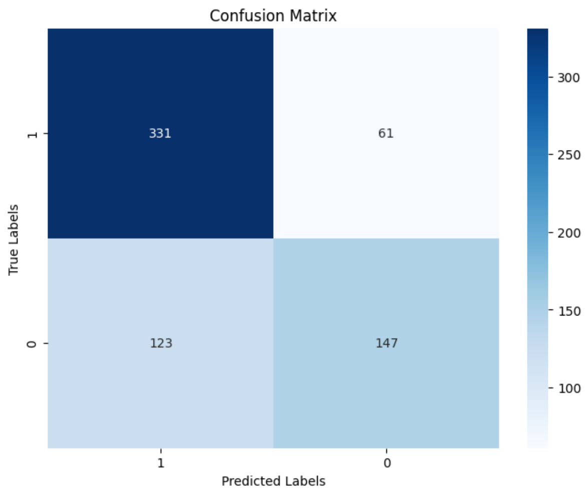

Figure 7. The Results of the Confusion Matrix of the KNN Model

Table 5. The Results of the KNN Model Evaluation

Accuracy	Precision	Recall	F1-Score
0.722	0.707	0.544	0.615

Table 5 shows that the results of the KNN model performance evaluation with the highest value are in the accuracy matrix = 0.722, then the precision value = 0.707, sequentially the F1-Score value = 0.615, and the recall value = 0.544. The value of the KNN model performance evaluation results reflects the extent to which the model can predict well against new data that has never been used for training. The accuracy value is 72% more, but with a deficiency of around 28%, the KNN model still needs to be improved; besides that, the F1-Score value is still small, which makes the KNN model unable to be implemented in real cases.

3.5 AdaBoost Model

The next model developed is the AdaBoost model, one of the machine learning algorithms. The AdaBoost algorithm is used to develop a model for determining whether a patient needs to be hospitalized or outpatient. Similar to the previous algorithms, the AdaBoost model is built by training data with the available parameters. After obtaining the AdaBoost model, the next step is to

This is an open access article under the CC BY-4.0 license

evaluate the model's performance using a confusion matrix to determine the extent to which the model can accurately predict new data. Visually, the results of the AdaBoost model performance evaluation using a confusion matrix are shown in Figure 8.

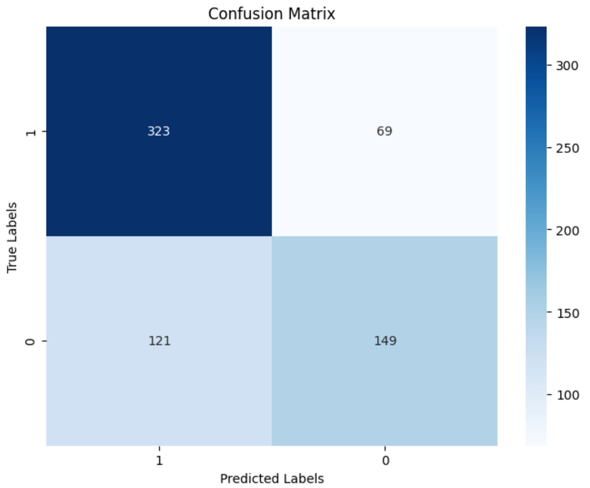

Figure 8. The Results of the Confusion Matrix of the AdaBoost Model

Based on Figure 8, it is known that the AdaBoost model can accurately predict 323 data as inpatients (1) and can accurately predict 149 data as outpatients (0). However, weaknesses are still visible in the AdaBoost model, with errors in predicting 121 data that should be outpatients (0) predicted as inpatients; in addition to the next error, there are 69 data that should be inpatients (1) predicted as outpatients. The results of the model performance evaluation with the confusion matrix can display the model's ability to predict new data so that it will be known how much data can be predicted accurately according to the original label. The AdaBoost model performance evaluation using several evaluation matrices was carried out to determine the model's performance in detail and comprehensively; the results of the model performance evaluation are shown in Table 6.

Table 6. The Results of the AdaBoost Model Evaluation

Accuracy	Precision	Recall	F1-Score
0.713	0.683	0.552	0.611

The detailed AdaBoost model performance evaluation results, as shown in Table 6, show that the model obtained an accuracy value = 0.713, and the lowest value was in recall = 0.552. In addition, the model obtained performance results with a precision value = 0.683 and an F1-Score value = 0.611. The results of the AdaBoost model evaluation show that the model is quite good in accuracy by obtaining 71%, but there are still 29% weaknesses that can be improved. To improve the model's accuracy, you can try to improve the data set, both in terms of quantity and features. In addition, of course, you can try various other methods in machine learning and can do hyperparameters with various methods that will be used. In this study, the focus is on how the methods in machine learning will be tried one by one to build a model by utilizing existing data.

3.6 MLPClassifier Model

The MLPClassifier method is one of the classification methods in machine learning. This study uses the MLPClassifier method to build a model by utilizing available data. Previously, data training has been carried out using various machine learning and neural network classification methods. The model performance evaluation was carried out after the model was obtained in the previous process. Similar to the previous process, when the MLPClassifier model has been obtained, the next step is to evaluate the performance, and the evaluation process begins by using a confusion matrix; the results are shown in Figure 9.

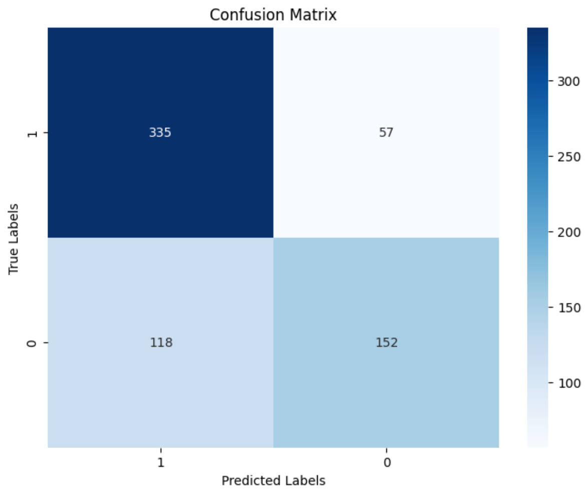

Figure 9. The Results of the Confusion Matrix of the MLPClassifier Model

The MLPClassifier model performance evaluation results are visualized, as shown in Figure 9. Based on these results, it can be seen that the MLPClassifier model can predict 335 data correctly with the predicted results of inpatient care (1), and 152 data can be predicted correctly as outpatient care (0). However, it is also seen that the MLPClassifier model still has weaknesses with errors in predicting 118 data as inpatient care (1) which should be predicted as outpatient care (0); in addition, there are still errors in predicting 57 data that should be inpatient care (1) predicted as outpatient care (0). These results show the extent to which the MLPClassifier model performs in predicting new data that has not been previously trained.

In addition to using a confusion matrix, the MLPClassifier model evaluation process needs to be carried out thoroughly to obtain results that are easy to understand and can be compared with other methods. Therefore, this study evaluated the MLPClassifier method for its performance using several matrices that are already popular for evaluation in classification cases. The results of the MLPClassifier model performance evaluation are shown in Table 7.

This is an open access article under the CC BY-4.0 license

Table 7. The Results of MLPClassifier Model Evaluation

Accuracy	Precision	Recall	F1-Score
0.736	0.727	0.563	0.635

The results of the MLPClassifier model evaluation have been carried out thoroughly, as shown in Table 7. The results show that the MLPClassifier model gets an accuracy value = 0.736. The accuracy value shows that a model will be able to predict new data; according to the evaluation results, the MLPClassifier model can be stated to be able to predict new data with an accuracy level of 74%. In addition, with the F1-Score value = 0.635, which calculates the precision and recall values results, the F1-Score value will be greatly influenced by how much the precision and recall values are. Based on the performance evaluation results, the MLPClassifier model still has weaknesses, with the model's inability to predict new data by 26%, so it can be a focus of the development process in the future.

3.7 Model Comparison

The results of the performance evaluation of the machine learning and neural network models have previously been explained in detail. Furthermore, a comparison is carried out to obtain the best model; the process of comparing models is focused on the accuracy matrix and F1-Score because both matrices can represent the overall model performance [33], [36]. The results of the model comparison are shown in Table 8.

Table 8. The Results of Model Comparison

Model	Accuracy	F1_Score
Neural Network	0.724	0.640
SVM	0.696	0.518
Decision Tree	0.662	0.564
KNN	0.722	0.615
AdaBoost	0.713	0.611
MLPClassifier	0.736	0.635

The best model is the model whose performance achieves the highest accuracy and F1-Score values [34], [37], [38], [39]. Based on the data displayed in Table 8, it can be seen that the performance of several models obtained accuracy values that were not much different, but in any case, with a slight difference, it can be used as a consideration in determining the best model. The comparison results showed that the MLPClassifier model was the best model, with an accuracy value = 0.736. The highest F1-Score was obtained by the Neural Network model with a value of 0.724. So in this study, the best model for determining patient care in Emergency Departments can use the machine learning model (MLPClassifier) and Neural Network. In this study, there are still several weaknesses that can be improved in the future, for example, the accuracy value; from several experiments using the machine learning model, it turns out that the results obtained an accuracy value that is not much different in the range of 74%, while for the F1-Score value, it is in the range of 63%.

MLPClassifier can outperform other models because MLPClassifier, as part of Neural Network, can capture more complex non-linear relationships compared to SVM, especially when the data has many features or dimensions. With an optimal architecture, the MLPClassifier tends to be more efficient for large datasets compared to SVM, which is often slow on large datasets due to the need to calculate the kernel matrix. In real-world data, such as clinical data, the class distribution is often imbalanced (for example, the number of positive cases is much smaller than the number of negative). Although MLPClassifier can handle this with certain adjustments (such as class weights), its performance still depends on data preprocessing. So that future research can focus on improving model performance, for example, by applying various approaches using deep learning [32], [40], [41], [42], [43].

4. Conclusions

This study produces a model built using machine learning methods (SVM, Decision Tree, KNN, AdaBoost, MLPClassifier) and neural networks. Various models obtained have been evaluated for performance using a confusion matrix and several matrices such as accuracy, precision, recall, and F1-Score. The results of the model performance evaluation are then compared using the accuracy matrix, and the F1-Score obtained the best model, namely the MLPClassifier model, which has an accuracy value = 0.736 and an F1-Score value = 0.635. In addition, the Neural Network model obtained an accuracy value = 0.724 and an F1-Score value = 0.640. So, the results of this study received the best model, namely the MLPClassifier and Neural Network models, because they could outperform other models. In this study, of course, there still needs to be an improvement in the performance of the existing model. The accuracy value is still around 74%, so model performance should be improved by trying a more sophisticated deep-learning approach for future research. Although MLPClassifier has shown the best results, testing with more complex deep learning architectures, such as Convolutional Neural Networks (CNN) for data that has spatial representation (such as medical images) or Recurrent Neural Networks (RNN) and Transformers for sequential data (e.g., patient medical history), can further improve performance.

References

[1] D. M. King, J. P. Vakkalanka, C. Junker, K. K. Harland, and A. S. Nugent, “Emergency Department Overcrowding Lowers Patient Satisfaction Scores,” *Academic Emergency Medicine*, vol. 28, no. 3, pp. 363–366, Mar. 2021, doi: 10.1111/acem.14046.

[2] P. McKenna, S. M. Heslin, P. Viccellio, W. K. Mallon, C. Hernandez, and E. J. Morley, “Emergency department and hospital crowding: Causes, consequences, and cures,” *Clin Exp Emerg Med*, vol. 6, no. 3, pp. 189–195, 2019, doi: 10.15441/ceem.18.022.

[3] A. A. Chamid, Widowati, and R. Kusumaningrum, “Multi-Label Text Classification on Indonesian User Reviews Using Semi-Supervised Graph Neural Networks,” *ICIC Express Letters*, vol. 17, no. 10, pp. 1075–1084, 2023, doi: 10.24507/icicel.17.10.1075.

[4] A. A. Chamid, Widowati, and R. Kusumaningrum, "Graph-Based Semi-Supervised Deep Learning for Indonesian Aspect-Based Sentiment Analysis," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 5, 2023, doi: 10.3390/bdcc7010005.

[5] N. Menteş, M. A. Çakmak, and M. E. Kurt, "Estimation of service length with the machine learning algorithms and neural networks for patients who receiving home health care," *Eval Program Plann*, vol. 100, no. September 2022, pp. 0–2, 2023, doi: 10.1016/j.evalprogplan.2023.102324.

[6] A. J. Hautala, B. Shavazipour, B. Afsar, M. P. Tulppo, and K. Miettinen, "Machine learning models in predicting health care costs in patients with a recent acute coronary syndrome: A prospective pilot study," *Cardiovasc Digit Health J*, vol. 4, no. 4, pp. 137–142, 2023, doi: 10.1016/j.cvdhj.2023.05.001.

[7] M. G. Siegel, M. J. Rossi, and J. H. Lubowitz, "Artificial Intelligence and Machine Learning May Resolve Health Care Information Overload," *Arthroscopy - Journal of Arthroscopic and Related Surgery*, vol. 40, no. 6, pp. 1721–1723, 2024, doi: 10.1016/j.arthro.2024.01.007.

[8] The Lancet Digital Health, "Machine learning to predict type 1 diabetes in children," *Lancet Digit Health*, vol. 6, no. 6, p. e374, 2024, doi: 10.1016/S2589-7500(24)00098-0.

[9] R. Daniel *et al.*, "Predicting type 1 diabetes in children using electronic health records in primary care in the UK: development and validation of a machine-learning algorithm," *Lancet Digit Health*, vol. 6, no. 6, pp. e386–e395, 2024, doi: 10.1016/S2589-7500(24)00050-5.

[10] R. Giddings *et al.*, "Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review," *Lancet Digit Health*, vol. 6, no. 2, pp. e131–e144, 2024, doi: 10.1016/S2589-7500(23)00241-8.

[11] M. J. Rho *et al.*, "Dr. Answer AI for prostate cancer: Intention to use, expected effects, performance, and concerns of urologists," *Prostate Int*, vol. 10, no. 1, pp. 38–44, 2022, doi: 10.1016/j.prnil.2021.09.001.

[12] R. O. Alabi, A. Almangush, M. Elmusrati, I. Leivo, and A. Mäkitie, "Measuring the Usability and Quality of Explanations of a Machine Learning Web-Based Tool for Oral Tongue Cancer Prognostication," *Int J Environ Res Public Health*, vol. 19, no. 14, 2022, doi: 10.3390/ijerph19148366.

[13] D. Benrimoh, M. Tanguay-Sela, K. Perlman, S. Israel, and J. Mehltretter, "Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician–patient interaction," *BJPsych Open*, vol. 7, no. 1, 2021.

[14] I. Ghanzouri *et al.*, "Performance and usability testing of an automated tool for detection of peripheral artery disease using electronic health records," *Sci Rep*, vol. 12, no. 1, pp. 1–11, 2022, doi: 10.1038/s41598-022-17180-5.

[15] A. J. Barda, C. M. Horvat, and H. Hochheiser, "A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare," *BMC Med Inform Decis Mak*, vol. 20, no. 1, pp. 1–16, 2020, doi: 10.1186/s12911-020-01276-x.

[16] J. C. Ginestra *et al.*, "Clinician Perception of a Machine Learning–Based Early Warning System Designed to Predict Severe Sepsis and Septic Shock," *Crit Care Med*, vol. 47, no. 11, pp. 1477–1484, 2019, doi: 10.1097/CCM.0000000000003803.

[17] S. Jauk, D. Kramer, A. Avian, A. Berghold, W. Leodolter, and S. Schulz, "Technology Acceptance of a Machine Learning Algorithm Predicting Delirium in a Clinical Setting: a Mixed-Methods Study," *J Med Syst*, vol. 45, no. 4, 2021, doi: 10.1007/s10916-021-01727-6.

[18] K. H. Bentley *et al.*, "Implementing Machine Learning Models for Suicide Risk Prediction in Clinical Practice: Focus Group Study With Hospital Providers," *JMIR Form Res*, vol. 6, no. 3, 2022, doi: 10.2196/30946.

[19] D. Gu, K. Su, and H. Zhao, "A case-based ensemble learning system for explainable breast cancer recurrence prediction," *Artif Intell Med*, vol. 107, no. August 2019, p. 101858, 2020, doi: 10.1016/j.artmed.2020.101858.

[20] N. Rahmayanti, H. Pradani, M. Pahlawan, and R. Vinarti, "Comparison of machine learning algorithms to classify fetal health using cardiotocogram data," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 162–171. doi: 10.1016/j.procs.2021.12.130.

[21] A. S. Zamani, A. H. A. Hashim, A. S. A. Shatat, M. M. Akhtar, M. Rizwanullah, and S. S. I. Mohamed, "Implementation of machine learning techniques with big data and IoT to create effective prediction models for health informatics," *Biomed Signal Process Control*, vol. 94, no. October 2023, p. 106247, 2024, doi: 10.1016/j.bspc.2024.106247.

[22] T. Halomoan Harahap *et al.*, "An artificial intelligence approach to predict infants' health status at birth," *Int J Med Inform*, vol. 183, no. January, p. 105338, 2024, doi: 10.1016/j.ijmedinf.2024.105338.

[23] A. Rahman, Z. Hossain, E. Kabir, and R. Rois, "Machine Learning Algorithm for Analysing Infant Mortality in Bangladesh," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2021, pp. 205–219. doi: 10.1007/978-3-030-90885-0_19.

[24] J. C. Ho, M. Sotoodeh, W. Zhang, R. L. Simpson, and V. S. Hertzberg, "An AdaBoost-based algorithm to detect hospital-acquired pressure injury in the presence of conflicting annotations," *Comput Biol Med*, vol. 168, no. June 2023, p. 107754, 2024, doi: 10.1016/j.compbiomed.2023.107754.

[25] A. A. Chamid, W. Widowati, and R. Kusumaningrum, "Text data labeling process for semi-supervised learning modeling," *12TH INTERNATIONAL SEMINAR ON NEW PARADIGM AND INNOVATION ON NATURAL SCIENCES AND ITS APPLICATIONS (12TH ISNPINSA): Contribution of Science and Technology in the Changing World*, vol. 3165, p. 030011, 2024, doi: 10.1063/5.0216320.

[26] A. A. Chamid, Widowati, and R. Kusumaningrum, "Labeling Consistency Test of Multi-Label Data for Aspect and Sentiment Classification Using the Cohen Kappa Method," *Ingénierie des Systèmes d'Information*, vol. 29, no. 1, pp. 161–167, 2024.

[27] Z. Cömert and A. F. Kocamaz, "Comparison of machine learning techniques for fetal heart rate classification," *Acta Phys Pol A*, vol. 132, no. 3, pp. 451–454, 2017, doi: 10.12693/APhysPolA.132.451.

[28] Mujiono Sadikin, "EHR Dataset for Patient Treatment Classification," 2020.

[29] Y. H. Shao, X. J. Lv, L. W. Huang, and L. Bai, "Twin SVM for conditional probability estimation in binary and multiclass classification," *Pattern Recognit*, vol. 136, Apr. 2023, doi: 10.1016/j.patcog.2022.109253.

[30] S. Kiran, G. R. Reddy, S. P. Girija, S. Venkatramulu, K. Dorthi, and V. Chandra Shekhar Rao, "A Gradient Boosted Decision Tree with Binary Spotted Hyena Optimizer for cardiovascular disease detection and classification," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100173.

[31] M. Chabbouh, S. Bechikh, C. C. Hung, and L. Ben Said, "Multi-objective evolution of oblique decision trees for imbalanced data binary classification," *Swarm Evol Comput*, vol. 49, pp. 1–22, Sep. 2019, doi: 10.1016/j.swevo.2019.05.005.

[32] D. Wang *et al.*, "Validation of a deep learning model for classification of pediatric pneumonia in Hong Kong," *Vaccine*, vol. 42, no. 26, Dec. 2024, doi: 10.1016/j.vaccine.2024.126370.

[33] M. Shirdel, M. Di Mauro, and A. Liotta, "Worthiness Benchmark: A novel concept for analyzing binary classification evaluation metrics," *Inf Sci (N Y)*, vol. 678, Sep. 2024, doi: 10.1016/j.ins.2024.120882.

[34] P. Naga Srinivasu, G. Jaya Lakshmi, S. Canavoy Narahari, J. Shafi, J. Choi, and M. Fazal Ijaz, "Enhancing medical image classification via federated learning and pre-trained model," *Egyptian Informatics Journal*, vol. 27, Sep. 2024, doi: 10.1016/j.eij.2024.100530.

[35] J. Mushava and M. Murray, "Flexible loss functions for binary classification in gradient-boosted decision trees: An application to credit scoring," *Expert Syst Appl*, vol. 238, Mar. 2024, doi: 10.1016/j.eswa.2023.121876.

[36] T. Baidawi and Solikhun, "A Comparison of Madaline and Perceptron Algorithms on Classification with Quantum Computing Approach," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 2, pp. 280–287, Apr. 2024, doi: 10.29207/resti.v8i2.5502.

[37] M. Pal *et al.*, "Ensemble approach of deep learning models for binary and multiclass classification of histopathological images

for breast cancer," *Pathol Res Pract*, vol. 263, p. 155644, Nov. 2024, doi: 10.1016/j.prp.2024.155644.

[38] G. S. Folli, M. H. C. Nascimento, B. P. O. Lovatti, W. Romão, and P. R. Filgueiras, "A generation of synthetic samples and artificial outliers via principal component analysis and evaluation of predictive capability in binary classification models," *Chemometrics and Intelligent Laboratory Systems*, vol. 251, Aug. 2024, doi: 10.1016/j.chemolab.2024.105154.

[39] N. D. A. Rahmah, B. Handoko, and A. A. Pravitasari, "Machine Learning Methods for Forecasting Intermittent Tin Ore Production," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 5, pp. 644–650, Oct. 2024, doi: 10.29207/resti.v8i5.5990.

[40] R. Qasrawi *et al.*, "Hybrid ensemble deep learning model for advancing breast cancer detection and classification in clinical applications," *Heliyon*, vol. 10, no. 19, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38374.

[41] A. El Kaffas *et al.*, "Development of a Deep Learning Model for Classification of Hepatic Steatosis from Clinical Standard Ultrasound," *Ultrasound Med Biol*, Nov. 2024, doi: 10.1016/j.ultrasmedbio.2024.09.020.

[42] Risfendra, G. Febri Ananda, and H. Setyawan, "Deep Learning-Based Waste Classification with Transfer Learning Using EfficientNet-B0 Model," *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 8, no. 4, pp. 535–541, Aug. 2024, doi: 10.29207/resti.v8i4.5875.

[43] S. Yang, P. Varghese, E. Stephenson, K. Tu, and J. Gronsbell, "Machine learning approaches for electronic health records phenotyping: a methodical review," *J Am Med Inform Assoc*, vol. 30, no. 2, pp. 367–381, Jan. 2023, doi: 10.1093/jamia/ocac216.