INTERNATIONAL JOURNAL OF
Telecommunications
AIR DEFENSE COLLEGE

A proposed heart disease diagnosis based on Deep learning.

Mohamed E. El-Bouridy[1], Amira S. El Batouty[2], Marwa Samara[3], Wael Abouelwafa Ahmed[4,] and Mohamed A. Massoud [5]

Citation: **E. El-Bouridy, Mohamed .; S. El Batouty, Amira .; Samara , Marwa.; Abouelwafa Ahmed , Wael .; A. Massoud , Mohamed .**
Title. Inter. Jour. of Telecommunications, ***IJT'2024***, *Vol. 04, Issue 02, pp. 01-11, 2024.*

Editor-in-Chief: Youssef Fayed.

Received: 27/05/2024.

Accepted: 15/07/2024.
Published: 15/07/2024.

Publisher's Note: The International Journal of Telecommunications, IJT, stays neutral regard-ing jurisdictional claims in published maps and institutional affiliations.

[1,2,3] Alexandria Higher Institute of Engineering &Technology, Alex;
[1]mohamed.ebouridy@aiet.edu.eg, [2]amira.elbatouty@aiet.edu.eg, [3]marwa.samara@aiet.edu.eg
[4,5] 3 Department of Biomedical, Minia University, Minia, 61519, Egypt.; [4]wael.wafa@minia.edu.eg,
[5]dr.massoud@mu.edu.eg; massoud300@yahoo.com,
* Correspondence: Mohamed.elbouridy@aiet.edu.eg; Tel.: (+203 01224950967 - 01501670002)

Abstract: One of the most influential factors in preserving a person's life is the late detection of heart disease, as cardiovascular disease is considered one of the biggest risks that lead to death. Cholesterol level, age, gender, as well as blood sugar level and heart rate are considered among the most influential factors in heart disease. The accurate diagnosis of all these diseases depends on the experience and skill of the treating physician. Many researchers have intended to use automated methods to diagnose diseases without relying on the expertise of doctors. In this research, the researchers present a proposal based on deep learning (DL) using the distinctive features of some factors affecting heart disease. Therefore, magnification techniques were used to diagnose whether the patient is at risk for cardiovascular disease. Bloody or not. The research resulted in progress, as accuracy in diagnosis reached 90.088%.

Keywords: Convolutional neural network (CNN)1, Sparse autoencoder2, deep learning (DL)3, and heart disease (HD)4.

1. Introduction

In [1], authors used a boosted decision tree algorithm (BDT) to determine associations between these cases, considering features of the patient and death Individuals were categorized as having a low or high chance of contracting the disease. Results estimated by the area under the curve were obtained with an accuracy of 88%. The authors of [2] achieved a 78% accuracy rate in the early identification of those at risk of developing heart failure by using electrocardiograms (ECGs) as CNN input. Authors have considered a few variables in a few research studies [3, 14, 15, 21, 22] to offer strategies that aid in the detection of cardiac illnesses, utilizing machine learning approaches to avoid issues resulting from statistical methods. In [4], Ali MM, and others describe how to choose the most suitable artificial intelligence classifications to achieve the greatest precision for these diagnostic situations. To examine their effectiveness and accuracy in predicting heart disease, a few supervised algorithms for machine learning were used. By examining the use of various machine learning strategies that rely on the stacked method produced by K-nearest neighbors (K-NN) with logistic regression analysis after K-NN classification utilizing the outputs of a voting classifier of prior techniques, an accuracy of 87.24% was reached in [5]. In 2022, Ghosh et al. presented the use of group algorithms, which showed good results in other diseases such as breast cancer, in addition to detecting heart failure [6], diabetic retinopathy [11], or hepatitis C [8], Note that many of these symptoms affect the functions of the heart, which may lead to death [7]. In [9,16] the authors used classical methods for predicting heart disease. However, by employing better methods, as in the instance of Faiz et al., the accuracy of these classifications can be increased. By suggesting an enhanced K-NN in [9], the authors were able to raise the diagnostic accuracy from the original K-NN by as much as 5.68%. Based on the Cleveland datasets with 297 records and applying the Hybrid Randomized Forest and Linear

Modeling (HRFLM) approach, the diagnostic accuracy was 88.7% in [16]. It is noted that all these research is very promising, but they need to increase the number of patients to generalize them. To train new algorithms and categorize a larger number of samples with a restricted number of those attributes, the authors in [10] built a new dataset based on certain well-known eleven datasets. Heart failure cases are an important topic for research because of the complex diagnostic procedures [12] that depend mainly on the expertise of doctors specializing in that field, which makes the search for methods for computer-assisted diagnosis very important, as shown in [17], where the time spent on accurate prediction has been reduced. Disease detection using data mining techniques. Regarding [13], the researchers integrated five classification model theories random forests (RF), Support Vector Machines (SVM), Naive Bayes Models (NBM), ANN, and Predictive Analysis (PA), to apply them to the diagnosis and prognosis of cardiovascular disease recurrence. Accuracy was achieved by using the repository's Cleveland and Hungarian sets of data. The RF technique produces 98.12%. Cardiovascular disease is one of the leading causes of death worldwide, accounting for more than seventeen million death yearly; it has been noticed that during the past few years, the number of persons suffering from heart disease has increased [16]. One of the most prevalent causes of death globally is cardiovascular disease (CVD) [18]. In developing nations, it is evident that heart disorders differ significantly in terms of the complications that accompany them and ultimately result in death [26]. Furthermore, it has been observed that heart failure-related mortality is rising in emerging nations and those with inadequate healthcare infrastructure [19]. This demonstrates the necessity of creating a technique that can result in the precise and timely assessment of a patient's risk of heart failure. Traditional machine learning models have demonstrated encouraging outcomes for this issue, as seen by numerous authors' attempts to blend various machine learning algorithms. For instance, a few multilayer perceptron (MLP) techniques were applied in [20], including generalized augmented regression, penalized regression, bootstrap pooled regression trees, multivariate adaptive regression lines, arithmetic mean, and Bayesian key-terms logistic regression with and without variable selection. It uses RF analysis, Classification Trees (CT), and the marginal risk of death assigned to each patient to predict the probability of death in heart failure patients within 30 days of discharge. These findings demonstrated that the combined models outperformed the regular models in terms of performance. Regarding [23], Peiris and colleagues proposed applying a variety of techniques, including SVM, ANN, BDT, K-NN, and stochastic gradient descent (SGD). It was observed that the results of SGD reached 87.69%, which is regarded as an excellent outcome in predicting the risk of heart disease. The article had a proposal for a technique that could identify heart illness with a respectable success rate, however it was only validated on a small sample of people. The idea of multilayer adaptive networks was applied in [24] to forecast the risk of heart failure, and the results revealed that it performed better than traditional neural networks and even hybrid and combination approaches that had been suggested earlier. The dataset was used in that article, and only two hundred ninety seven patients were assessed, making it a tiny sample size. A broad spectrum of disorders affecting the heart, blood vessels, and the human body's circulatory and distribution system are together referred to as "cardiovascular disease" [25]. Particularly in underdeveloped nations, the large range of heart problems causes many complications that can result in mortality [26].

Now the desired goal of this research and the steps taken to achieve it will be reviewed, which is to obtain results with high percentages for classifying heart diseases, which leads to a high success rate in early detection of the disease. Using the evaluated data set, two secondary goals were achieved:

- Use very few features to create a new problem classification methodology.
- Secondly. Make use of CNN special qualities to enhance current feature augmentation methods.

A classifier based on CNN and a sparse autoencoder was presented to process the data for analysis to meet the study goals. The achievement of up to 90% accuracy is regarded as a significant breakthrough in heart disease risk assessment.

This study will be given in the following three sections: The traditional techniques for comparing the outcomes and the structures employed are described in the methodological section. The data set utilized, the experiments, and the results will all be covered in detail in the Experiments and Results section. Lastly, a discussion of the conclusions will be provided.

2. Methodology:

The results obtained from the proposed architecture have been compared with several well-known classical approaches such as: K-NN, AdaBoost, Gaussian Naive Bayes (GNB), MLP, BDT, RF, and XG Boost.

The k-NN method is a lazy machine learning algorithm that learns from test data, not from training data. It uses vectors of dimension p as:

$$xi = (x1i, x2i,...,xpi) \in X \quad (1)$$

and calculates the distance between stored and new vectors. The data is classified with the class that repeats the most in the selected vectors, using the Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{p}(x_{ri} - x_{rj})^2} \quad (2)$$

AdaBoost is a method that combines multiple weak classifiers to create a robust classifier. It assigns greater weights to poorly classified data and less to well classified data, focusing on bad cases. The algorithm uses decision trees as the base method and involves initializing weights, training weak classifiers, and combining them into a strong classifier.

The GNB theorem is used in classification to predict continuous data, converting the data into a frequency table and probability table, and applying Bayes' theorem to select the class with the highest posterior probability.

A MLP is a neural network with input, output, and intermediate layers, with a linear activation function, connecting neurons in each layer to learn complex information from input data.

A BDT is a model where each node is labeled with an input feature, allowing for decision-making between multiple options. Arcs are unions between nodes, leading to subordinate decision nodes on different input features. Leaf nodes are labeled with a class or probability distribution, containing the final selected class.

RF is a learning algorithm that combines decision trees, with each tree defining the number of data and classifier variables. The input variables are chosen for each node, and the best partition is calculated. The tree is iterated, and the one with the highest number of indices is used as a predictor.

XG Boost classifier compares results with the proposed classifier, focusing on reducing losses and error in iterations, unlike AdaBoost which assigns more weight to misclassified samples.

3. Proposed Diagnostic Technique

The proposed diagnostic technique will be addressed through a detailed presentation and proposal of a sparse autoencoder, then exposing data classification using convolutional neural networks, and then presenting a multi-task artificial neural network proposal.

3. 1 Refactoring and Feature Augmentation: sparse autoencoder (SAE)

Less neurons make up the latent space of a typical autoencoder than the input and output layers, allowing for a smaller representation of the feature vector. The latent space has a greater number of neurons than both the input and output layers when utilizing XG Boost. The network uses only a few neurons at a time because of adding an L1 regularization term to the latent space. We can enhance the quantity of data attributes and conduct different analyses of them by using this network structure. A typical architecture of SAE is shown in Figure 1. Consequently, the sparse autoencoder divides the neural network's decoder section and ascertains which part of the encoder accesses the latent space once it has been trained to reconstruct the input. As a result, the encoder separates the original N features into M features (with N<M) by maximizing the features of the input data. This process makes it possible to save both the hidden additional information and all the original data information.

3. 2 Classification of Data: convolutional neural networks (CNN).

CNN, which extracts complicated features from 2D data like photos, can be fed this type of data. The filters (kernels) are what allow this information to be extracted. The filter weights are changed during the training phase to apply a precise feature map for every class.

Each convolutional layer in the CNN architecture needs to be followed by a pooling layer for the computation to be heavy. Thus, the number of network parameters and overfitting are minimized by the pooling layers. Max-pooling, the most popular kind of pooling, operates by establishing the maximum value for every window. For CNN to classify the features that the kernels have retrieved, its final layers need to be dense layers. Figure 2 shows the conventional CNN.

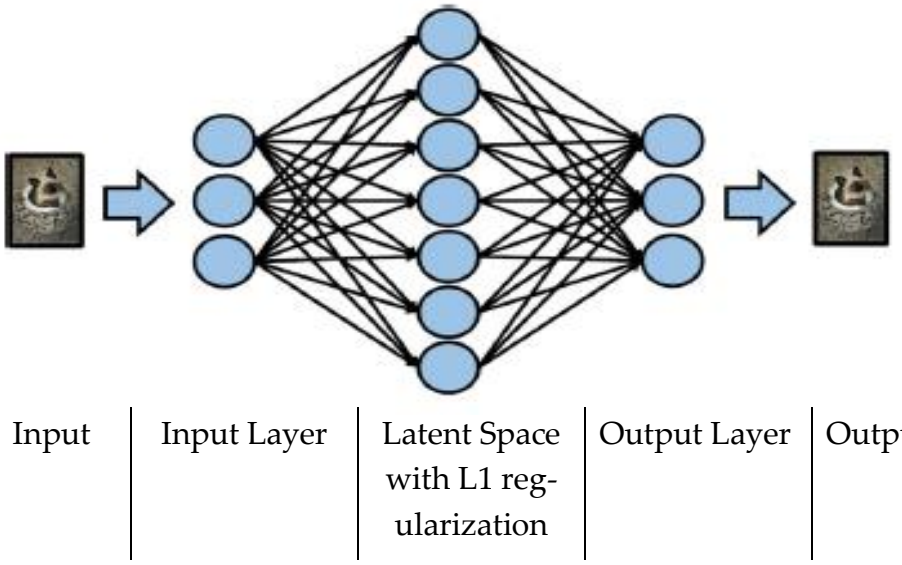

Figure 1. Typical SAE architecture. Applying the L1 regularization term to the latent space.

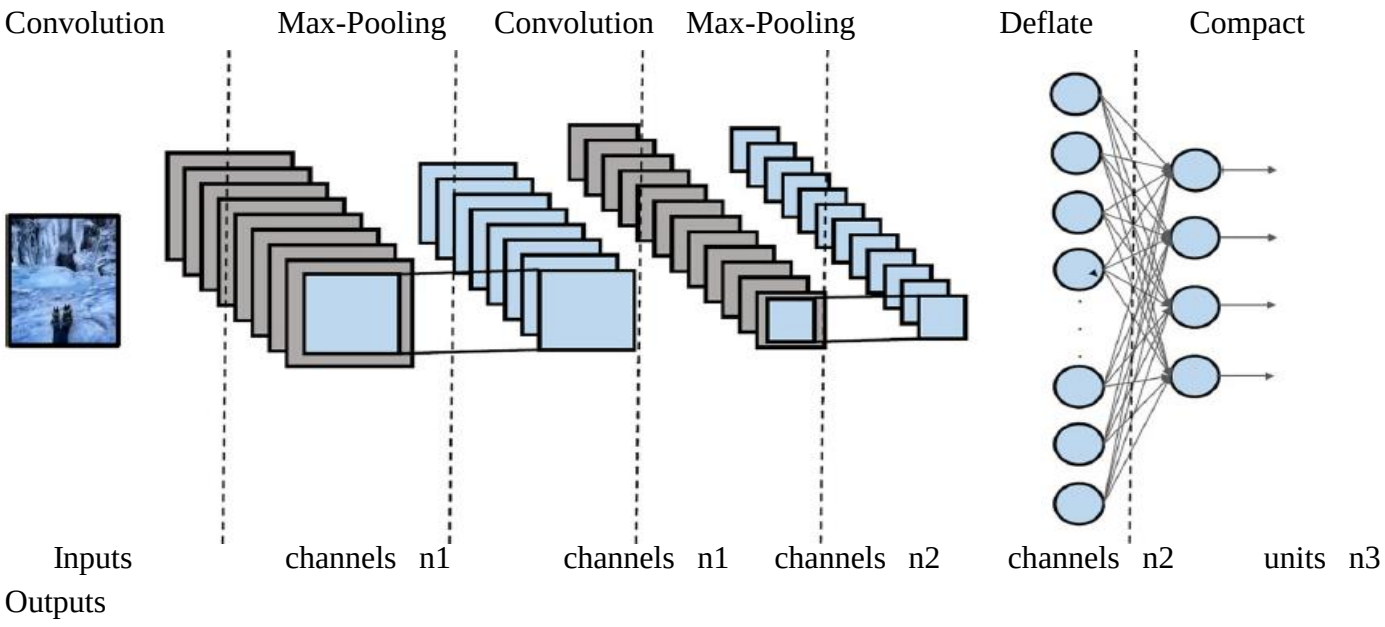

Figure 2. Conventional CNN.

3.3 The Proposed: the multitask artificial neural network.

A structure for the permuted network was proposed to implement two tasks, namely using the SAE to enhance the features of all the data. Additionally, A classifier that was learned using SAE is linked to the latent space. As a result, the evaluation was carried out utilizing two distinct classifiers (to join the SAE): the CNN in Figure 4 and the conventional MLP in Figure 3.

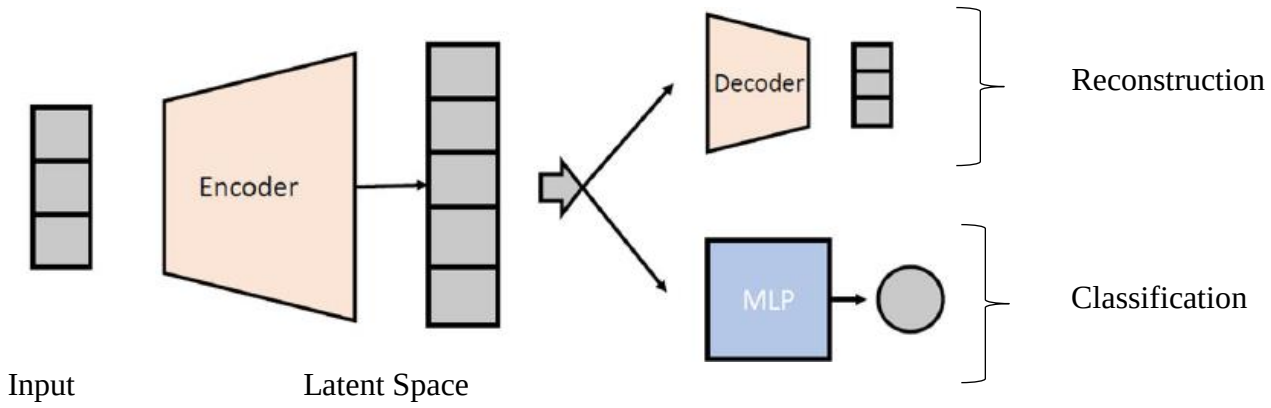

Figure 3. A multitask neural network consisting of an MLP classifier and a Sparse autoencoder**.**

Three different strategies for handling data with few features are shown in Figure 5. In the first technique, the feature augmentation procedure is used to expand the number of features by using the complete training data set as input. The classification model then uses the updated data set, which has more features than the initial set. and whether a positive class is present in the sample is then ascertained. In the second method, the small data set is supplied straight into the model, bypassing the feature improvement stage. The third suggested method is based on the simultaneous training of a classifier network that uses the largest dataset as input to predict the final class and a data augmenting NN known as Sparse Autoencoder. As a result, by training the suggested model with this method, the feature augmenting process becomes trainable and can be enhanced to produce predictions that are more accurate as the data is processed.

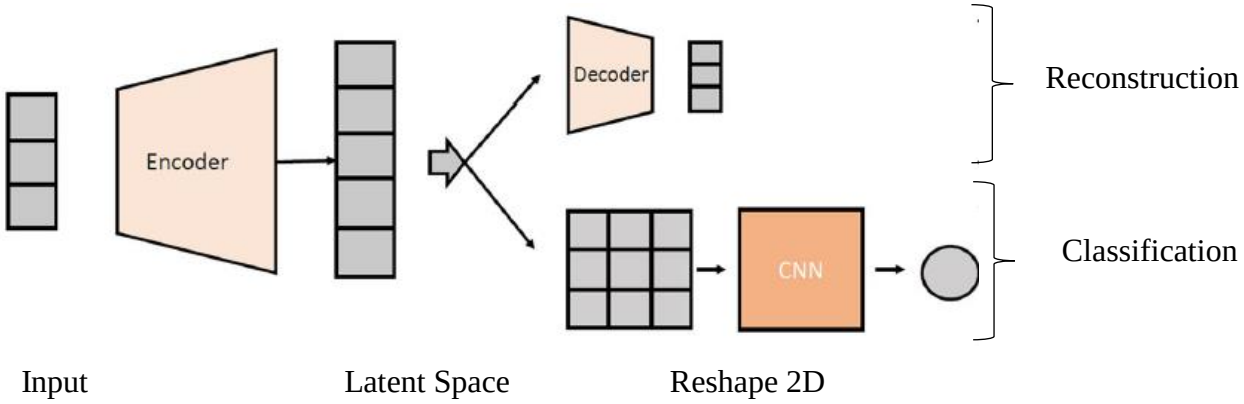

Figure 4. Multi-task neural network configuration consisting of a sparse Autoencoder and a CNN classifier.

4. Experiences and results.

4.1 Description of the data set.

Eleven clinical characteristics made up the dataset: sexuality, age of the patient, resting arterial blood pressure in millimeters, type of pain in the chest (atypical, nonanginal, or asymptomatic pain), fasting blood sugar (value 1 if fasting > 120 mg/dL, 0 otherwise), serum cholesterol in milliliters, allowing ECG results (that can be normal, ST if a patient has ST-T defects, or LVH if the patients indicates cardiac hypertrophy), exercise-induced chest pain, which may be categorized as yes or no, maximum age (a value in years based on depressive disorders), highest rate of heart rate (a numbers value across sixty people and 202), and finally, the ST-

segment sloping of maximum heart rate (greater, straight bottom). The output category, which can be either 0 (normal) or 1 (the heart illness), is shown in Column 12. This set of data was created by merging several datasets Cleveland, Hungarian, Swiss, Long Beach, and Staling that were previously available separately but had not been integrated. The last set of data on cardiovascular disease has 918 samples, with 410 samples representing the healthy class and 508 samples representing the class of problems with the heart. Each class has a comparable number of cases [10]. This eliminated the need for strategies to address uneven class sizes.

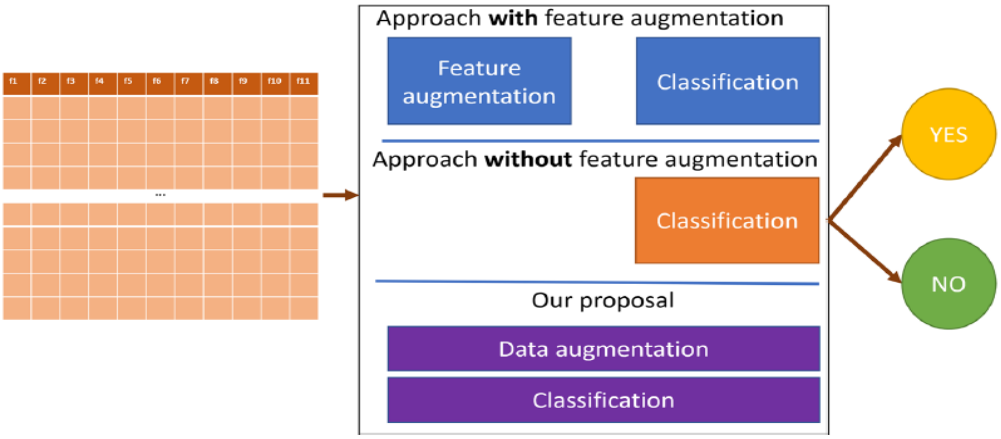

Figure 5. Three distinct strategies for handling data with few features.

4.2 A Testing setting.

Initially, data pre-processing step was implemented to tidy up and take out more valuable data from the data set by dividing it into three groups on each of the age groups and the rest BP feature, the cholesterol feature, and the rapid coding technique on three features: Chest Pain Type, Resting ECG, and ST Slope. AND Angina, related to sex and exercise using a label coding program. In this process, the classification of age data was modified to represent the three age groups: young people, adults, and older adults. Also, the rest of the BP feature has been modified to three new columns for low BP, medium BP, and high BP. The cholesterol characteristic has also been converted into three classification columns that determine risks: low, medium, and high. At the end of that process, the data set consisted of 24 features.

K-fold cross-validation was carried out with 10-fold cross-validation, where each model was assessed by a thorough hyperparameter network search, to prevent randomness in all tests. In the results, the best average value is presented for 10 times with the superior parameter configuration. For feature augmentation, the neural network architecture consists of a discrete, discrete encoder, which is trained at the same time. Therefore, two different formulations were performed. The first will serve the work of the MLP, in contrast, the second trains a 2D neural network by converting the intrinsic region into a 2D matrix. Due to its quick improvement time, stability in reconciliation graduation, and capacity to operate with sparse gradients—which take neural network performance into account—ADAM improvement was used for training in both scenarios. For the classification subnet, binary cross-entropy was the loss function that was employed, and for the decoder, mean squared error. One way to define binary cross-entropy is using Equation (3).

$$-\big(y\log(p)+(1-y)(1-p)\big) \tag{3}$$

Because binary cross-entropy is equal to employing maximum likelihood when developing a model, it was used. Thus, the following is a definition for the mean square error:

$$\sum_{i=1}^{D}(x_i-y_i)^2 \tag{4}$$

It's employed in autoencoder problems since it penalizes heavily for significant errors and has great potential for feature reconstruction. We evaluated multiple latent space sizes to see how important this parameter was for the final classification.

To obtain the results of classification, the used data was divided into two groups. The first group represents 70% of the total data used and was used in the training process. The second group represents 30% of the total data used and was used in the testing process.

4.3 Final result.

An investigation of the traditional machine learning techniques previously described in methodology was carried out to compare the theory of action put forward in this work with other theories of action. Consequently, a grid search was used to identify the most effective super learning techniques for each technique. The average 10-fold cross-validation using the optimal setup is displayed in Figure 6.

The neural network's performance outcomes employing several classical theories are displayed in Figure (6). The chart illustrates that the best accuracy of 86.281% was obtained by utilizing MLP theory. It was followed in accuracy by the random forest with a rate of 86.279%, then the AdaBoost group method with a very close rate of 86.276%. While K-NN theory obtained a score of 85.512%, followed by XG-Boost theory with a score of 85.191%. Then the GNB theorem with a rate of 84.637%. The worst accuracy when using D-tree theory was 78.978Because the neural networks performed well, training was carried out by fusing MLP with feature improvement via a patchy automatic encoder. The findings obtained utilizing various latent space sizes are displayed in Figure (7). where a cross-validation coefficient of 10 is employed to represent each average accuracy number.

The accuracy of the classification increased to 89.543% while a latent space containing 100 features was used rather than the initial eleven features. These findings also showed that training the MLP and SAE simultaneously with a multi-task neural network improved categorization. We find that this method not only works better at identifying cardiac issues, but it also makes it possible to add new features to the dataset. Subsequently, a 2D CNN was used in place of the MLP classifier in a new series of studies.

It should be noted that the CNN method handles structured data by employing SAE to reorganize the data to add the best possible visual representation by combining existing characteristics to create new ones. A comparison of these findings for various latent space sizes is presented in Figure 8.

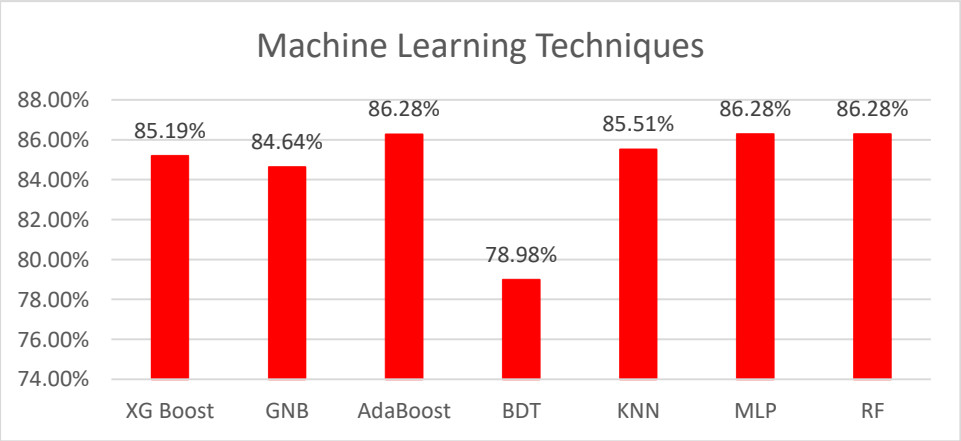

Figure 6. The outcomes attained with traditional machine learning techniques.

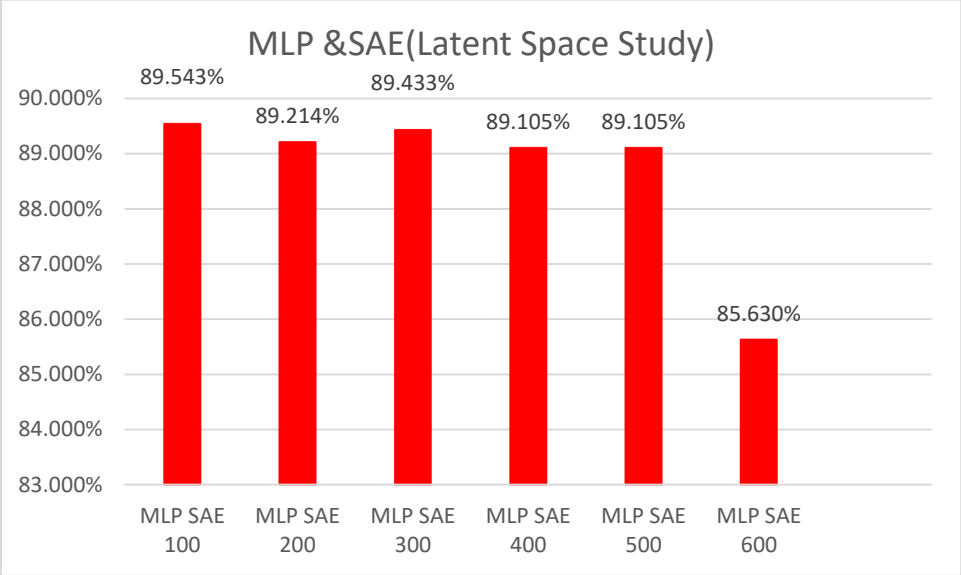

Figure 7. The obtained results of classification when using the combinations architecture of MLP with sparse autoencoder for feature augmentation.

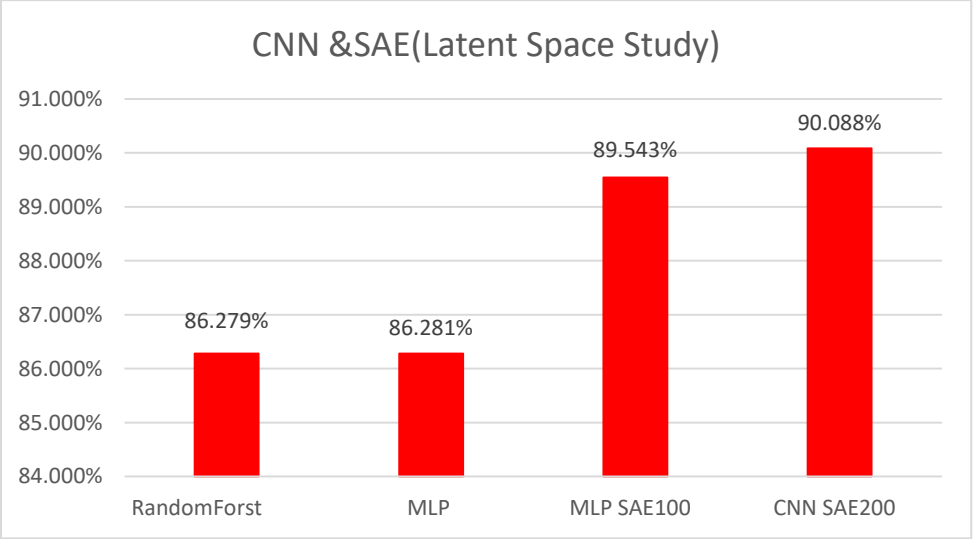

Figure 8. The results were obtained using a sparse autoencoder and an architecture combining CNN for classification and feature augmentation.

After conducting several experiments with a latent region containing 200 more features, the best result was obtained with a precision of 90.088%. Studying Figure 9, we see the improvement in the proposal presented regarding the Random Forest model and the vanilla MLP model. Feature augmentation techniques improved the results obtained on the original dataset. Furthermore, by re-ranking new features utilizing a combined strategy of SAE and 2D CNN, the accuracy can be marginally boosted.

Two tests, the Kolmogorov-Smirnov test, and the t-test for independent samples, were employed to confirm that the experimental findings produced using the suggested strategy are superior to those obtained by more conventional techniques.

The methods employed were divided into two groups to test the accuracy results: the first group consisted of conventional machine learning and MLP approaches, while the second group included the methods proposed and reported in this research, which are MLP and CNN employing SAE. After conducting these tests, the results showed that, with a significance of $p < 0.001$, the accuracy acquired by the suggested approach ($M = 88.99$,

and SD = 1.13) improved in compared to the accuracy achieved through conventional techniques (M = 84.73, and SD = 2.61). For a t-test, t (17) = 4.97, and p<0.001, is the result.

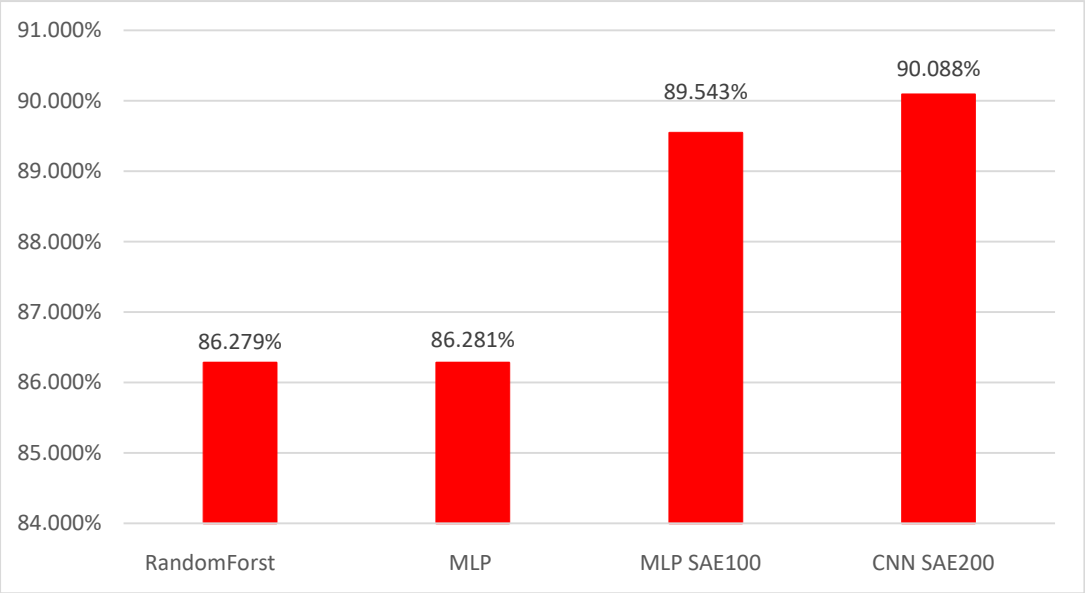

Figure 9. The suggested multi-task neural networks are compared to Random Forest and traditional MLP models.

5. Conclusions.

Using patient records from five different centers as a dataset, this study provides deep learning-based approaches for feature augmentation and ensemble classification tasks to address cardiovascular disease prediction. There were 918 samples in this collection, and each sample had just eleven clinical features. Next, a new architectural strategy that combines a sparse autoencoder and a convolutional classifier is proposed. Since the dataset only has eleven features, feature improvement was done using a spare autoencoder to extract new features. CNN was trained by resampling it into a 2D matrix using the retrieved characteristics. These two steps, which combine an SAE and a classifier (MLP or CNN) in a complex network, were added to improve the ability to extract features while taking the classifier information into account for the backpropagation algorithm's feedback. We observe that CNN outperforms MLP by 0.6% accuracy when SAE is concurrently trained. After a thorough investigation of the number of neurons in the latent space of the sparse autoencoder, which represents new characteristics, we find that 200 is the ideal size.

The results of this study also suggested that having more neurons is not always a good thing. An accuracy of 90.088% was achieved with this method, which is regarded as a 4.4% improvement over the outcomes of traditional classifiers (MLP or RF) trained on the same dataset and under the same conditions.

To achieve clinical integration, we must consider many of the challenges that we will encounter when implementing this integration, such as potential considerations for integrating these results into clinical workflow. As well as data privacy, regulatory compliance, and seamless integration with current health care information systems, which are recommended to be studied extensively in Upcoming research.

The hardware resources: HP ZBook Studio G7 Mobile Workstation, Operating system: Windows 10 Pro 64-bit, Processor: Intel Core i9-10885H, RAM: 32 GB DDR4, and Storage: 1 TB PCIe Gen 3 x4 NVMe TLC SSD. The computational time required for training and inference about 30 mints.

References

1. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, UreyMAet al (2020) Improving risk prediction in heart failure using machine learning. Eur J Heart Fail 22(1):139–147. https://doi.org/10.1002/EJHF.1628

2. Akbilgic O, Butler L, Karabayir I, Chang P, Kitzman D, Alonso A et al (2021) Artificial intelligence applied to ecg improves heart failure prediction accuracy. J Am Coll Cardiol 77(18):3045. https://doi.org/10.1016/S0735-1097(21)04400-4

3. Albert KF, John R, Divyang P, Saleem T, Kevin MT, Carolyn JP et al (2019) Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines. Circ Arrhythmia Electrophysiol, vol 12(7). https://doi.org/10.1161/CIRCEP.119.007316

4. Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA (2021) Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Comput Biol Med 136:104672. https://doi.org/10.1016/J.COMPBIOMED.2021.104672

5. Araujo M, Pope L, Still S, Yannone C (2021) Prediction of heart disease with machine learning techniques. Graduate Res, Kennesaw State Un

6. Dalal S, Onyema EM, Kumar P, Maryann DC, Roselyn AO, Obichili MI (2022) A hybrid machine learning model for timely prediction of breast cancer. Int J Model Simul Sci Comput 0(0):2341023. https://doi.org/10.1142/S1793962323410234

7. Diwakar M, Tripathi A, Joshi K, Memoria M, Singh P, Kumar N (2021) Latest trends on heart disease prediction using machine learning and image fusion. Mater Today: Proc 37(Part 2):3213–3218. https://doi.org/10.1016/J.MATPR.2020.09.078

8. Edeh MO, Dalal S, Dhaou IB, Agubosim CC, Umoke CC, Richard-Nnabu NE et al (2022) Artificial intelligence-based ensemble learning model for prediction of hepatitis C disease. Front Public Health 10:892371

9. Faiayaz Waris S, Koteeswaran S (2021) Heart disease early prediction using a novel machine learning method called improved K-means neighbor classifier in python. Mater Today: Proc, https://doi.org/10.1016/J.MATPR.2021.01.570

10. Fedesoriano Heart failure prediction dataset kaggle. Available from https://www.kaggle.com/ fedesoriano/heart-failure-prediction. Accessed 12 September 2022

11. Ghouali S, Onyema E, Guellil M, Wajid MA, Clare O, Cherifi W et al (2022) Artificial intelligencebased teleopthalmology application for diagnosis of diabetics' retinopathy. IEEE Open J Eng Med Biol, pp 1–11. https://doi.org/10.1109/OJEMB.2022.3192780

12. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ et al (2014) Heart disease and stroke statistics—2014 update. Circulation, vol 129(3). https://doi.org/10.1161/01.CIR.0000441139. 02102.80

13. Jan M, Awan AA, Khalid MS, Nisar S (2018) Ensemble approach for developing a smart heart disease prediction system using classification algorithms. Res Rep Clin Cardiol 9:33–45. https://doi.org/10.2147/RRCC.S172035

14. Khajehali N, Khajehali Z, Tarokh MJ (2021) The prediction of mortality influential variables in an intensive care unit: a case study. Personal Ubiquit Comput, https://doi.org/10.1007/s00779 021-01540-5

15. Kim YJ, Saqlian M, Lee JY (2022) Deep learning–based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining. Personal Ubiquit Comput 26(2):259–267. https://doi.org/10.1007/ s00779-019-01248-7

16. Kondababu A, Siddhartha V, Kumar BB, Penumutchi B (2021) A comparative study on machine learning based heart disease prediction. Mater Today: Proc. https://doi.org/10.1016/J.MATPR.2021.01.475

17. Krishnaiah V, Narsimha G, Chandra NS (2016) heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. Int J Comput Appl 136(2):975–8887

18. Maini E, Venkateswarlu B, Maini B, Marwaha D (2021) Machine learning–based heart disease prediction system for Indian population: an exploratory study done in South India. Med J Armed Forces India 77(3):302–311. https://doi.org/10.1016/J.MJAFI.2020.10.013

19. Muzammal M, Talat R, Sodhro AH, Pirbhulal S (2020) A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. Inf Fusion 53:155–164. https://doi.org/10.1016/ J.INFFUS.2019.06.021

20. Negassa A, Ahmed S, Zolty R, Patel SR (2021) Prediction model using machine learning for mortality in patients with heart failure. Am J Cardiol 153:86–93. https://doi.org/10.1016/J.AMJCARD.2021.05.044

21. Olsen CR, Mentz RJ, Anstrom KJ, Page D, Patel PA (2020) Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. Am Heart J 229:1–17. https://doi.org/10. 1016/J.AHJ.2020.07.009

22. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J (2015) Using EHRs and machine learning for heart failure survival analysis. Stud Health Technol Inform 216:40

23. Pires IM, Marques G, Garcia NM, Ponciano V (2020) Machine learning for the evaluation of the presence of heart disease. Procedia Comput Sci 177:432–437. https://doi.org/10.1016/J.PROCS.2020.10.058

24. Samuel OW, Yang B, Geng Y, Asogbon MG, Pirbhulal S, Mzurikwao D et al (2020) A new technique for the prediction of heart failure risk driven by hierarchical neighborhood component-based learning and adaptive multi-layer networks. Future Gener Comput Syst 110:781–794. https://doi.org/10.1016/J. FUTURE.2019.10.034

25. Soni J, Ansari U, Sharma D, Soni S (2011) Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int J Comput Appl 17(8):43–48. https://doi.org/10.5120/2237-2860

26. Yang H, Garibaldi JM (2015) A hybrid model for automatic identification of risk factors for heart disease. J Biomed Inform 58:S171–S182. https://doi.org/10.1016/J.JBI.2015.09.006