**JOURNAL OF CRITICAL REVIEWS**

ISSN- 2394-5125 VOL 06, ISSUE 06, 2019

# DIAGNOSING AND PREDICTING PULMONARY HYPERTENSION USING MACHINE LEARNING METHODS

**Ahmed I. Taloba[1], Osama R. Shahin[2], Rasha M. Abd El-Aziz[3]**

1Information System Department, Faculty of Computers and Information, Assiut University, Assiut, Egypt, taloba@aun.edu.eg

2Physics and Mathematics Department, Faculty of Engineering, Mataria, Helwan University, Egypt,osama.r.shaheen@gmail.com

3Computer Science Department, Faculty of Science, Assiut University, Assiut, Egypt, rashamahmoud@aun.edu.eg

**Abstract:**

The risk classification of young hypertensive patients stays difficult. Pulmonary hypertension was once thought to be an incurable disease. Theunderstanding of diseases impacting the circulation of pulmonary has changed dramatically during the last two decades. There have been 30 randomised controlled trials (RCTs) investigating therapies and operation for patients with chronic thromboembolic pulmonary hypertension is being established. Improvements in imaging have allowed for additional comprehensive patient valuation, yet pulmonary hypertension remains a life-threatening illness with a two-year gap between the onset of symptoms and detection. For predictive tasks, machine learning (ML) is a strong technique in finding as well as organising multiple useful factors. There are no techniques for early detection and screening of pulmonary hypertension (PH), which is serious to managing development as well as minimizing related death. To date, pulmonary hypertension (PH) has had a high percentage of missing diagnoses, but there has been little progress in developing a quick, easy, and efficient technique to test for the disease. The project's goal is to create a machine learning strategy for detectingprobable abnormalities in chest radiographs that could indicate PH in patients who are suspected of having it. There have been no techniques for early detection andscreening of pulmonary hypertension (PH), which is crucial in addressing development and minimizing related death. A decision tree method was created and verified as a machine learning approach for detecting PH utilizing electrocardiography (ECG). The findings of the experiments reveal that the decision support system has the highest level of precision, demonstrating its utility in the healing of pulmonary hypertension.

*Keywords:* Pulmonary hypertension, Machine learning, Decision tree, SVM, ECG.

## 1. Introduction:

Machine learning (ML) is a word used in the development of artificial intelligence research to study powerful algorithms that can be used to resolve issues unsupervised or supervised [1]. Machine learning (ML) methods were seen to undertake well in the areas of treatment, medical phenotyping [2], clinical risk, and clinical outcome estimation in the past few years

ISSN- 2394-5125 VOL 06, ISSUE 06, 2019

because they give an indiscriminate criterion for choosing variables based on their impact on outcomes and the classifier methods could process huge amounts of data with complicated interplay [3]. ML is a cutting-edge computing technique that may enhance the processing of healthcare data and make medical decisions automatically [4]. Yet, just several studies of applying machine learning methods to forecast the outcome of people with hypertension have now been published to date, and the cohorts analysed have primarily been middle-aged or older patients[5]. Machine learning methods have evolved as very efficient computational models for identifying patterns in large medical datasets with many variables, making it easier to build a model for data-driven forecasting or categorization [6]. Pulmonary hypertension can be predicted and classified using machine learning and current data mining approaches. In clinical research, classification is commonly employed. The precise classification of pulmonary hypertension allows for a more accurate diagnosis of patients. Boosting is a versatile strategy that combines successful rules to produce highly precise predictions. Machine learning approaches are varying decision trees that mix boosting algorithms and decision tree in producing easy-to-understand categorization rules.

At cardiac catheterization, pulmonary hypertension is defined as a mean PAP of 25 mm Hg or higher. The original clinical categorization in 1973 emerged from a World Health Organization sponsored international conference after an epidemic related to the use of the appetite suppressant, aminorex fumarate[7]. Primary pulmonary hypertension was classified as having pulmonary artery vasculopathy. PH is characterized by a rise in average pulmonary arterial pressure (PAP) at rest as measured by right heart catheterization (RHC) [8]. It is a pathologic illness that can affect numerous chronic conditions. Pulmonary hypertension is extremely diversified, which range from uncommon and often severe increases in PAP in pulmonary arterial hypertension (PAH) to more frequent and often moderate increases of pressure in respiratory and cardiac disease [7]. Pulmonary hypertension (PH) is a chronic, progressing condition that causes decreased dyspnea, exercise adaptation, mortality and right heart failure [9]. PH is now classified into 5 categories with identical pathological and clinical features: pulmonary arterial hypertension (PAH), chronic thromboembolic pulmonary hypertension (CTEPH), pulmonary hypertension with unknown/multifactorial mechanisms, pulmonary hypertension due to left heart disease (PH-LHD) and pulmonary hypertension due to lung disease and/or hypoxia (PH-lung)[10].

Identifying the disease, determining the type of pulmonary hypertension (which dictates medication), assessing prognosis, and monitoring response to therapy are all important components of pulmonary hypertension management [11].At right heart catheterization (RHC), PH is described as an average pulmonary arterial pressure (mPAP) of $\geq$ 25 mmHg [12]. The present standard solution for PH is right-sided heart catheterization (RHC), with just a resting average PAP of 25 mm Hg or greater deemed diagnostic for PH [13]. For a non-invasive original estimate of pulmonary arterial pressure, echocardiography is routinely utilised [14]. The standard for evaluating right ventricular (RV) structure and functioning is magnetic resonance imaging (MRI), which provides higher resolving power and precision than echocardiography, as well as a high-level degree of consistency of quantitative measures of ventricular morphological features and functions[15].An alternate cause of dyspnea may be discovered using electrocardiography, chest radiography, and pulmonary function tests. In a large IPAH registry, electrocardiography revealed right heart strain and chest radiography revealed significant pulmonary arteries or cardiomegaly in 80-90 percent of cases. A normal electrocardiogram and chest radiograph, on the other hand, cannot rule out the diagnosis in clinical practise.

Electrocardiography (ECG), chest radiography, and echocardiography are frequently used to diagnose PH. Even though most patients with PH have ECG irregularities corresponding to right heart overflow at diagnoses, the conventional ECG is insufficient for monitoring, as well as chest radiographs are not specific for differentiating PH from other heart illnesses [16]. In terms of PH monitoring, echocardiography is the gold standard. Although international recommendations recommend echocardiographic values when diagnosing PH, echocardiography is an expensive, time-consuming, and difficult-to-access tool for early identification. The most prevalent symptom of PAH is increasing dyspnea. Patients with right ventricular dysfunction may experience exertional dizziness and syncope. Late in the disease, oedema and ascites develop[17]. Exertional myocardial hypoperfusion in the setting of right ventricular hypertrophy, pulmonary artery compression of the left major coronary artery, or ischemic heart disease can cause anginal chest discomfort. The breathless patient is methodically evaluated and high-risk categories are screened to identify pulmonary hypertension. In the absence of symptoms or signs of left heart disease or respiratory, unexplained progressive exertional breathlessness is suggestive. For high-risk groups, screening techniques exist. Patients with systemic sclerosis are screened every year at rheumatology centres, and potential liver transplant receivers are frequently tested for portopulmonary hypertension. After a pulmonary embolism, persistent dyspneamust alert the doctor to the risk of CTEPH. The following fig.1 shows the categories of machine learning method.

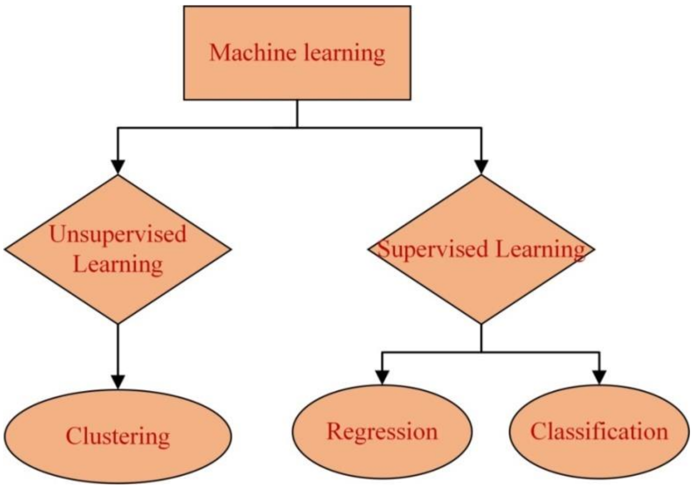


**Fig.1 Categories of Machine learning techniques**

Machine learning approaches is being increasingly popular for analysing, predicting, and classifying medical information in past few years [18]. Machine learning and data mining methods could be used to identify hypertension in its initial stages. By collecting significant data from patient data, these approaches may efficiently detect pulmonary hypertension at an

ISSN- 2394-5125 VOL 06, ISSUE 06, 2019

early stage [19]. We looked into how machine learning methods could help with echocardiographic pulmonary hypertension (PH) diagnosis. In this research, a machine learning method called decision tree classifier is utilised to identify and forecast pulmonary hypertension. The following is how the rest of the paper is organized: In Section 2, a systematic literature review was conducted to discover new a study on the topic of PH. Section 3 depictsa detailed explanation of the proposed method. Section 4 discusses the result and discussion in detail. Section 5 concludes the study with a conclusion.

## 2. Related works:

This paper [20]investigates the possibility of using machine learning methods to forecast results in younger patients and performance was evaluated to one of the currently utilised techniques in clinical practice. In 508 people with HP who'd been allowed to treat at a tertiary care centre, baseline medical evidence as well as a composite endpoint—which included new-onset atrial fibrillation/atrial flutter, all-cause death, coronary artery revascularization, acute myocardial infarction, new-onset heart failure, peripheral artery revascularization, new-onset stroke, sustained ventricular tachycardia/ventricular fibrillation. At the 33-month follow-up assessment, the ML method was constructed using recursive feature reduction, severe riseimproving, as well as 10-fold cross-validation. Using the ML technique, an 11-variable combination was shown to be the most useful to predict results. For evaluating the medical outcome of young people with hypertension, the ML methods were compared to Cox regression and outperformed the recalibrated Framingham Risk Score model. In this experiment, unfortunately, the ML model only forecasted endpoint occurrences and did not account for censorship or follow-up duration, nor did it describe the relationship among predictors and results.

Deep learning methods were employed in the current work [21] to categorize anomalies in chest radiographs that suggested PH with extreme precision and generalisation. The technique might give a non-invasive and simple way to screen people diagnosed as having PH if it is proven prospectively in healthcare situations. The total number of photographs in this technique was 762. After doing 8-fold cross-validation on the 641 photos chosen for training, we opted to set the education rate at 0.0008 based on the top validation result.Lastly, the models are trained using all of the validation andpre-training data, then tested on external and internal testing data to categorise the pictures as showing PH symptoms or normal according to the area under the receiver operating characteristic curve. Following that, the regression approach was utilised to forecast PASP using the three deep learning techniques. When applying DL techniques in differentiating PH from healthy based on X-rays, the top method got 0.970 in the internal test and somewhat fell and got 0.967 in the external test. However, more research is needed to see if these results are feasible in a prospective clinical context.

The goal of this research [22] is to see if the machine learning approaches combined with image-based PH measures may enhance the diagnosis precision of MRI in PH. Within 48 hours, 72 patients with alleged PH who visited a referral centre undertook MRI and RHC. 57 patients were found to have PH, while 15 had none. To test the diagnostic value of the combination, researchers combined the total of structural as well as functional cardiovascular and cardiac markers obtained from two mathematical methods into a classification system. The findings of a decision support approach that combined computation-derived measures indicating hemodynamic variations in the pulmonary vasculature with measurements of right ventricular architecture as well as functioning were confirmed utilizing leave-one-out cross-validation, which offers a technique to noninvasively identify PH with higherprecision (92%).In patients with suspected PH, the highest diagnostic precision of these MRI-based design variables can minimize the necessity for RHC. Even though it's indicative for referring patients to a PH referral centre, in which such methods can be used to eliminate cardiac catheterization, the major drawbacks of this single-centre, retrospective analysis are the limited model size of the analysing cohort and the disparity among high- and low-risk people.

An artificial intelligence method was created and verified for diagnosing PH utilizing electrocardiography in this approach [23]. Data from patients diagnosed from 2 hospitals were used in this historical cohort study. Patients at one hospital are split into extraction and internal validation datasets, while those in the other have been solely included in external validation data.Using demographic dataand 12-lead ECG signal from the dataset, an Artificial Intelligenceprocedure-based ensemble neural network is created. The identification of PH served as the study's endpoint. Using a sensitivity map, the interpretable AI system determined that area had the greatest impact on decision-making. The area under the receiver operating characteristic curve of the Artificial Intelligence system in identifying PH is 0.902 as well as0.859 during external and internal validation, respectively. During the follow-up phase, those people who the AI classified as getting a greater risk than those in the less risk group had a considerably better risk of getting PH. Utilizing single-lead ECGs as well as 12-lead data, the AI model achieved great precision in PH diagnosis.

The goal of this research [24] is to see whether the machine learning of three-dimensional forms of systolic cardiac motion might forecast survival of patients and causes of right ventricular failure in PH. Bythe mean follow-up of 4.0 years, 256 people with recently identifiedPH received 6-minute walk testing, right-sided heart catheterization, and cardiac MRI. The most highly predicted patterns of systolic motion are identified using supervised principal components analysis. The findings suggest using machine learning to forecast patient results in PH using complicated motion characteristics collected from cardiac MR imaging.The inclusion of every treatment regimen can restrict applicability in certain sets;however, the study reveals that the procedures are efficient all over a wide range of diseases and therapies. To eliminate preference in the classification of cause of death, the article's censoring and the endpoint is limited to all-cause mortality, although the performances of indicators would be influenced by differences in treatment and those chosen for the operation. The segmentation method's precision compares favourably to that of earlier studies; yet, the ambiguity in end-diastolic and end-systolic segmentation may spread to the ambiguity in displacement estimate.

Leveraging the potential of machine learning has showed substantial improvements in a variety of lung disorders, according to this study. These could be used to enhance pulmonary hypertension radiological evaluation and clinical care. Quantitative imaging, particularly, can result in a data-driven decision-making procedure that better assesses clinical, physiological, genetic, and radiological data. It could aid with the field's current clinical and radiological challenges, allowing for better phenotyping and evaluation to make treatment options. Imaging-based biomarkers, if confirmed, could be utilised as outcomes in clinical trials to evaluate response to therapy. The expanding application of artificial intelligence (AI) would help minimize errors, boost productivity, and allow anaccurate medicine method for pulmonary hypertension. With an emphasis on lung disease, this state-of-the-art article presents current QCT discoveries and possible applications in patients with PH. To summarise, it's envisaged that the use of AI techniques to CT imaging will "come to the aid" for clinicians are managing patients with PH and related lung illness by offering mechanistic insights and enhanced phenotyping, hence facilitating much-needed therapeutic trials.

## 3. Proposed Methodology:

This research offers the framework depicted in Fig. 2 for assessing the risk of pulmonary hypertension using a combination of models. Data collection, pre-processing, trainingand testing with the models, and comparing the findings and prediction of pulmonary hypertension are the six parts of this methodology. The parts have been discussed in the following sections.

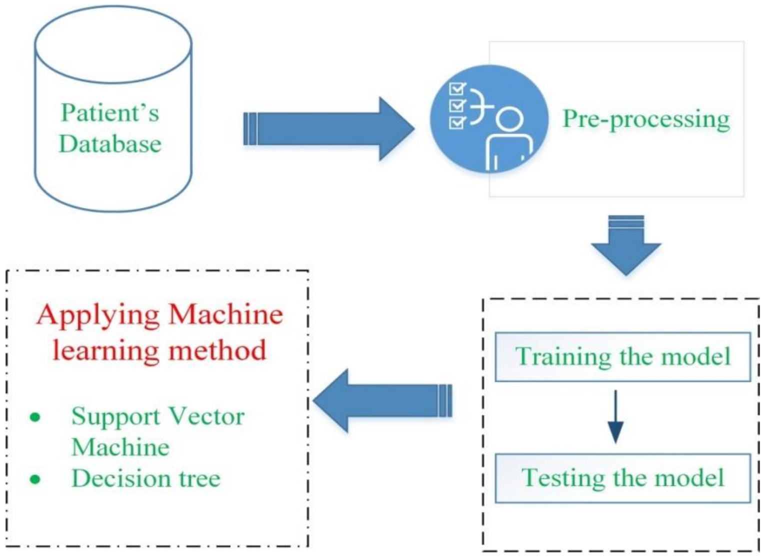


**Fig. 2 Proposed Framework**

### 3.1 Database collection:

The outcomes of a retrospective analysis on ECG tests as well as the outcomes of RHC done at the Clinic for Cardiology and Pulmonology are presented in this paper [12]. The research was carried out as a database search confined to ECG and RHC data, as permitted by local ethics committees and in compliance with the modified Declaration of Helsinki. Before the data was viewed, it was thoroughly anonymised. The following are the 11 characteristics that are taken into account:age, sex,Weight (kg), Height (cm), BMI (kg/m2), Pulse rate (PULSE), Haemoglobin (g/dL), Hypertension (HTN), pulmonary arterial pressure (mmHg), Pulmonary heart disease (PHD), Diabetes mellitus (DM). The database contains 500 patient records in total.

### 3.2 Data Pre-processing:

The data from the Clinic for Cardiology and Pulmonology is extracted in a standard manner during this phase. The next stage is to convert raw data that includes removing missing variables, normalising the data, and eliminating outliers. Six tuples contain missing attributes out of the 500 official records. These have been left out of the analysis. SVM manually centred data points at their mean and scaled them to have a unit standard deviation. The data sets for decision trees do not need to be changed. Fig. 3 shows the Basic measures of Pulmonary Arterial Hypertension.

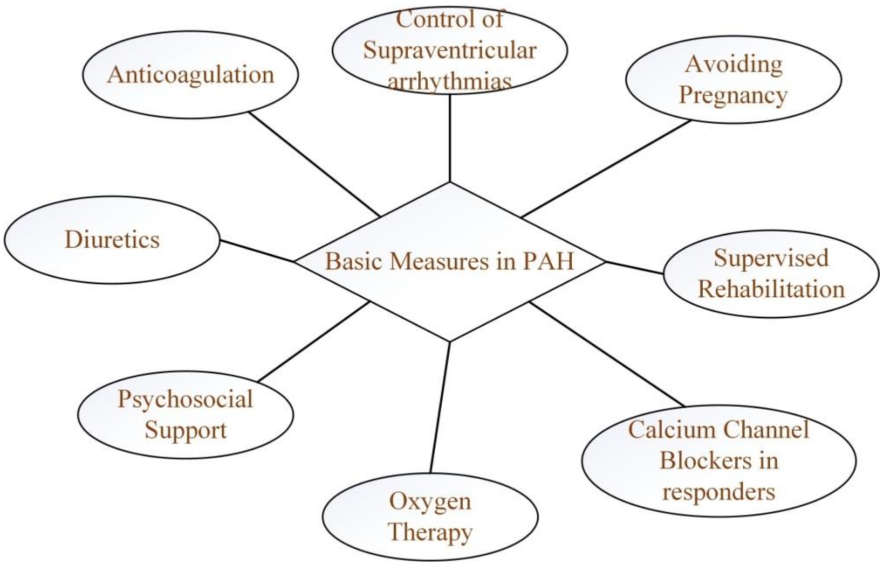


**Fig.3 Basic measures of PAH**

### 3.3 Training the Models

Various approaches were used to train each of the three models. A node splitting criterion is necessary for decision trees. A split thatseparates the data into distinct categories is the best. Purity is a measure used to evaluate the likelihood of a split. The purest split is one thatdivides an attribute into 2 distinct classes. As a result, it is recommended that it is used as the splitting criterion for decision trees, while alternative standards are used instead, resulting in varying accuracy. The most effective training approach for support vector machines is K-fold cross-validationthatdivides the data into k blocks and averages the output of the blocks.This approach trains the data with all of the tuples before testing it with one of the blocks. For training, 10-fold cross-validation is commonly utilised. The models' effectiveness and estimation of the general error were evaluated using 10-fold cross-validation, which is a popular method in computer science. The model requires ten rounds of training and testing, with various training and testing sets every time. The outcome was approved as the mean of the ten test findings.

### 3.4 Testing the Models :

### 3.4.1 Support Vector Machine:

Support vector machine (SVM) is commonly utilised in classification and regression issues because they are computationally strong tools for supervised learning. Statistical learning theory (SLT) and Bayesian considerations motivate the systematic method. The SVM classifier's main goal is to determine the best dividing hyperplane among negative and positive samples. The best hyperplane is the one that provides the most margin among training examples that are closest to the hyperplane. Text categorization, Handwritten digits classification, facial recognition, particle identification, as well as bioinformatics are just a few of the real-world challenges that SVM classifiers have effectively solved. In many situations, SVMs have stronger generalisation than many other classification methods, however, this enhanced generalisation comes at a cost. SVMs are much slower than other methods in the testing phase.

In regression and classification analysis, a support vector machine is a type of method which is utilized to evaluate informationas well as discover patterns. If your data has exactly2 classes, thena support vector machine (SVM) is utilized. The optimum hyperplane which divides every data point of one class from those of the other class is found by an SVM. SVM is a mathematical function-based approach that can be used to describe complex, real-world issues. SVM works well with data sets with a lot of characteristics. The training data is mapped into kernel space using Support Vector Machines. To mention a few, it containspolynomial kernels, linearkernels, Multilayer Perceptron kernels, quadratic kernels, Radial Basis Function kernels, and so on.Least squares, sequential minimal optimization, and Quadratic programming are all methods that are used to build SVM.

Procedure selection andKernel are problematic aspects of SVM to get correct so that your algorithm isnot overly optimistic or pessimistic. Given the enormous number of instances and characteristics in the gathered information, it is doubtfulwhether the linear kernel orRBF must be utilized. Even though the connection among class labels and features is nonlinear, the RBF kernel may not enhance performance because of the vast number of characteristics. Both kernels should be tested, with the more effective one being chosen in the end.

### 3.4.2 Decision Trees :

A decision tree is a method that can be used to help with decision-making and ML. It has a tree-like appearance. A group of internal nodes, a root node, as well as a group of terminal nodes known as leaves, make up the tree. Every node in a decision tree generates a binary choice that divides one or more classes from the other classes. A decision tree is a technique that predicts a reaction to data using categorization or regression. If the features are grouped, classification is used, and if the data is continuous, regression is used.

One of the most used data mining methods is the decision tree. To analyse the data, follow the path from the root to the leaf node. As mentioned in the training session, decision trees must be formed using a purity index that splits the nodes. Eachtuple is reviewed down the decision tree using the gathered data, resulting in a positive or negative assessment of pulmonary hypertension. These are evaluated to the dataset's original decision parameter to verify for false positives or negatives, yielding the accurateness, specificity, and sensitivity. The decision tree algorithm is the widely used classification algorithm since they are simple to know and apply. Figure 4 shows the decision tree classifying the data set.

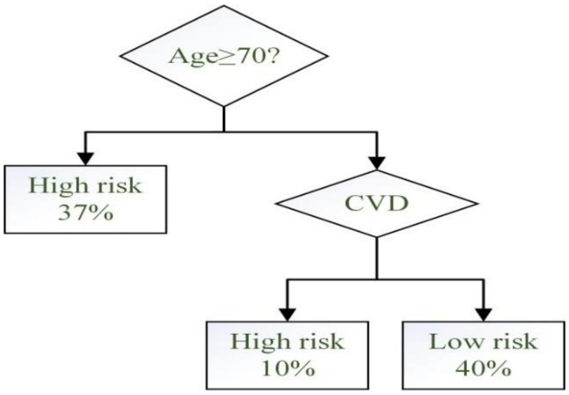


**Fig. 4 Decision tree classifying pulmonary hypertension data**

**Table 1: Attributes used**

| Sl.No | Attributes |
|---|---|
| 1 | Sex |
| 2 | Age |
| 3 | Weight |
| 4 | Height |
| 5 | BMI |
| 6 | Pulse rate |
| 7 | Haemoglobin |
| 8 | Hypertension |
| 9 | Pulmonary arterial pressure |
| 10 | Pulmonary heart disease |
| 11 | Diabetes mellitus |

A patient record with 11 features is used in our research. Attributes used in our research are shown in Table 1. Fig. 4 depicts the flowchart for the proposed framework.

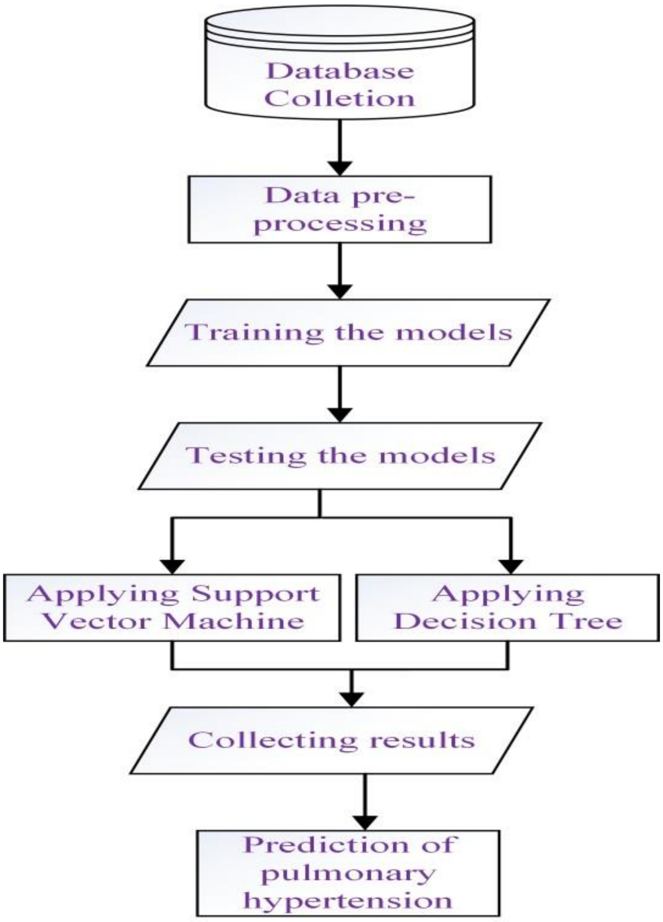


**Fig.5 Flowchart for the proposed methodology**

Fig.5 depicts the flowchart for the proposed methodology. The following algorithm shows the procedure to diagnose and predict pulmonary hypertension.

| **Algorithm: Proposed Machine learning algorithm** |
|---|
| **Input:** Data collection using ECG |
| **Output:** Diagnosing and predicting Pulmonary hypertension |
| Step 1: Data is collected by ECG signals |
| Step 2: Collected data is then pre-processed |
| Step3: The data model is trained |
| Step 4: To test the models the Machine learning methods are implemented |
| Step 5: The result is collected to diagnose and predict the PH |

### 3.5 Statistical analysis:

SPSS (Statistical Package for the Social Sciences) is used to conduct the statistical analysis. The mean ± standard deviation of numerical data was calculated, whereas a categorical variable was reported as percentages (%). All variables are tested first to assess statistical differences among samples. The following Table 2 shows the characteristics of different patients.

**Table 2: Characteristics of various patients' mean and standard deviation**

| Characteristics | | Total |
|---|---|---|
| Age | | 30.79±6.08 |
| Sex | Male | 378 (75.0%) |
| | Female | 122 (25.0%) |
| Height, cm | | 170±10 |
| Weight, kg | | 76±14 |
| BMI, kg/m2 | | 26.24±4.43 |
| Pulse rate (PULSE) | | 97.8 ± 16.1 |
| Haemoglobin, g/dL | | 12.6±2.1 |
| Hypertension (HTN) | | 319 (63.8%) |
| Pulmonary arterial pressure, mmHg | | 79±59 |
| Pulmonary heart disease (PHD) | | 79 (15.8%) |
| Diabetes mellitus (DM) | | 69 (13.8%) |

## 4. Result and Discussion:

The proposed model is put to the test with the information received. The ECG waves are used to identify pulmonary hypertension. These signals are fed into the system, which then undergoes a pre-processing step to eliminate any noise. The sensitivity, specificity, and accuracy of the findings obtained using the SVM and decision tree algorithms will be examined. The AUC and ROC curve was used to assess the accuracy of the proposed diagnostic metrics. Table 3 shows that the best technique had an average AUC of 0.69 on the training data set, as determined by 10-fold cross-validation.The test data shows an AUC of 0.92.

**Table 3: The AUC was calculated by 10-fold cross-validation on both the training and testing data**

| **Training Set** | | | | **Testing set** | |
|---|---|---|---|---|---|
| **Dataset** | **AUC** | **Dataset** | **AUC** | **Dataset** | **AUC** |
| 1 | 0.59 | 6 | 0.69 | Test set | 0.92 |
| 2 | 0.63 | 7 | 0.76 | | |
| 3 | 0.69 | 8 | 0.71 | | |
| 4 | 0.74 | 9 | 0.72 | | |
| 5 | 0.72 | 10 | 0.60 | | |
| **Average AUC is 0.69** | | | | | |

Results from the sensitivity analysis conducted byECG Characteristics are shown in Table 4.The results are statistically significant at the $p < 0.05$ level.

ISSN- 2394-5125 VOL 06, ISSUE 06, 2019

**Table 4: Cardiac Phenotypes according to Patient**

| Echocardiographic Characteristics | Patient | p-value |
|---|---|---|
| LVMi, g/m$^2$ | 99.3 ± 23.8 | <0.0001 |
| LVEDV, ml | 107.7 ± 20.6 | <0.0001 |
| LVEDVi, ml/m$^2$ | 54.4 ± 8.9 | <0.0001 |
| LVESV, ml | 42.2 ± 9.7 | <0.0001 |
| LVESVi, ml/m$^2$ | 21.4 ± 4.7 | <0.0001 |
| LVEF, % | 61.5 ± 5.4 | <0.0001 |
| E/A ratio | 1.6 ± 0.6 | <0.0001 |

Using the training and testing set the predicted and actual risk of occurring pulmonary hypertension using machine learning is shown in Fig. 6. By the figure the predicted risk is higher than the actual risk.

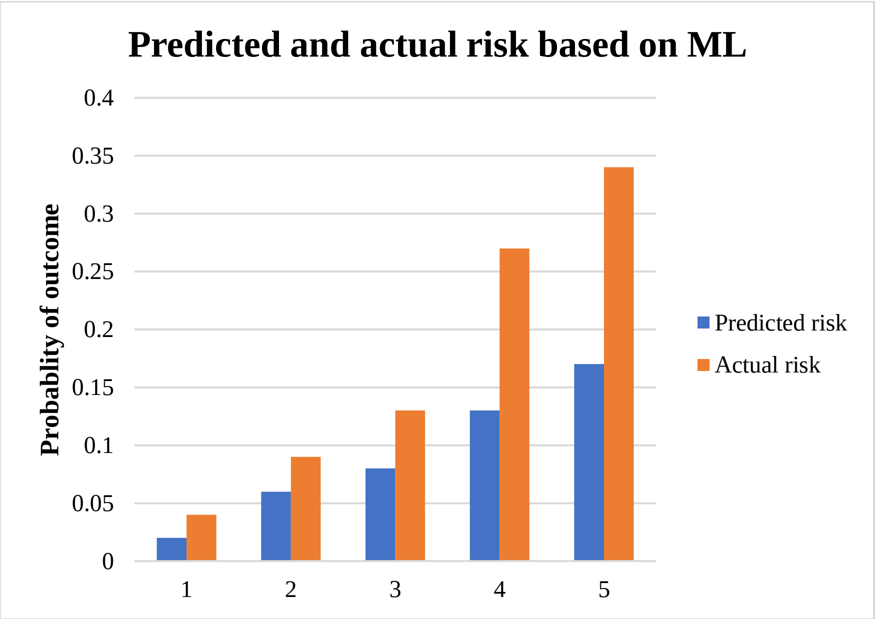


**Fig. 6 Predicted and actual risk**

All approaches produce AUC values greater than 0.78, and all confidence ranges are nearly identical. When compared to existing approaches, the proposed method achieved a higher AUC value. Fig.7 shows the AUC and ROC curve for the proposed method.

ISSN- 2394-5125 VOL 06, ISSUE 06, 2019

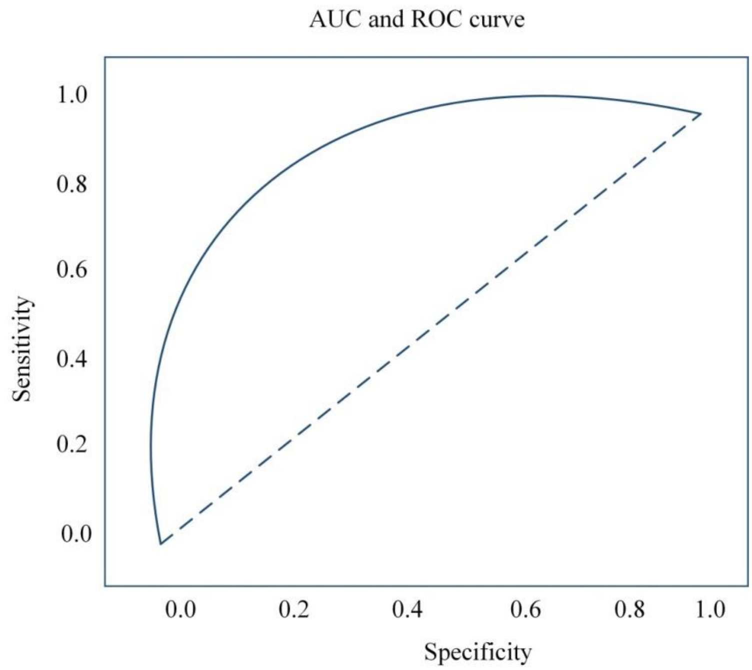


**Fig. 7 AUC and ROC curve**

Table 5 illustrates the detailed performance metrics of the proposed method using the 10-fold cross-validation method.

**Table 5: Performance metrics of the method**

| **Proposed Method** | |
|---|---|
| **Parameters** | **Percentage (%)** |
| Sensitivity | 0.97 |
| Specificity | 0.95 |
| Accuracy | 0.98 |
| AUC curve | 0.92 |

Fig. 8depicts the testing time accuracy of the proposed Support vector Machine and Decision tree. Among them,the Decision tree shows higher accuracy when compared to SVM.

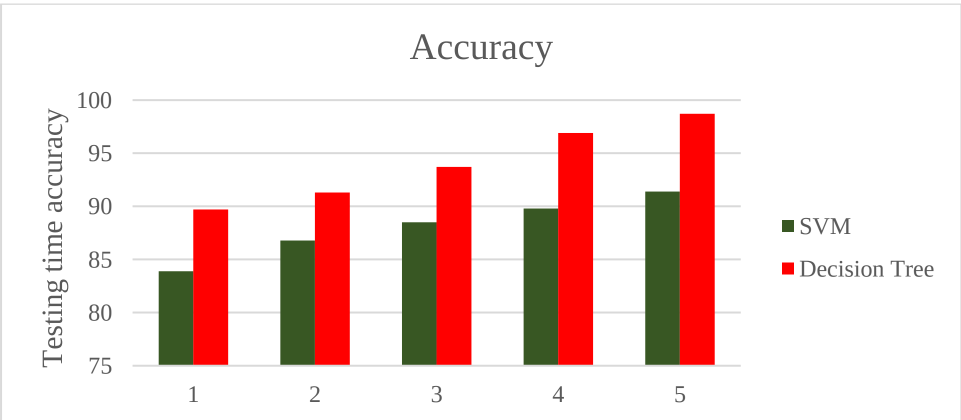


**Fig. 8 Testing time accuracy**

## 5. Conclusion:

A delayed or missing diagnosis of PH can have serious consequences. We investigated the relevance of ML-based statistics in the difficult clinical prediction of PH since ML algorithms can be simply integrated into echocardiographic devices. As a result of the literature analysis, more complicated and combinational models are needed to improve the accuracy of predicting the early development of pulmonary hypertension. This research provides a method for accurately predicting pulmonary hypertension using a mix of support vector machines and decision trees. Among the two the decision tree method performs well with a higher accuracy rate. Following that, this work presented a comparative analysis of the various outcomes, including sensitivity, specificity, and accuracy. Furthermore, the most efficient and heavily weighed model may be discovered. The next phase of work is developing the system utilising the aforementioned approaches, as well as teaching and testing it. Thus, the proposed method predicts pulmonary hypertension using a machine learning technique with an accuracy of 98.7%.

## References:

[1] A. J. Tajik, “Machine Learning for Echocardiographic Imaging,” *J. Am. Coll. Cardiol.*, vol. 68, no. 21, pp. 2296–2298, Nov. 2016, doi: 10.1016/j.jacc.2016.09.915.

[2] L. Ernande *et al.*, “Clinical Implications of Echocardiographic Phenotypes of Patients With Diabetes Mellitus,” *J. Am. Coll. Cardiol.*, vol. 70, no. 14, pp. 1704–1716, Oct. 2017, doi: 10.1016/j.jacc.2017.07.792.

[3] S. Basu, S. Raghavan, D. J. Wexler, and S. A. Berkowitz, “Characteristics Associated With Decreased or Increased Mortality Risk From Glycemic Therapy Among Patients With Type 2 Diabetes and High Cardiovascular Risk: Machine Learning Analysis of the ACCORD Trial,” *Diabetes Care*, vol. 41, no. 3, pp. 604–612, Mar. 2018, doi: 10.2337/dc17-2252.

[4] X. Liu, K. Chen, T. Wu, D. Weidman, F. Lure, and J. Li, “Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer’s disease,” *Transl. Res.*, vol. 194, pp. 56–67, Apr. 2018, doi: 10.1016/j.trsl.2018.01.001.

[5] R. C. Lacson, B. Baker, H. Suresh, K. Andriole, P. Szolovits, and E. Lacson, “Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients,” *Clin. Kidney J.*, vol. 12, no. 2, pp. 206–212, Apr. 2019, doi: 10.1093/ckj/sfy049.

[6] A. M. Darcy, A. K. Louie, and L. W. Roberts, “Machine Learning and the Profession of Medicine,” *JAMA*, vol. 315, no. 6, p. 551, Feb. 2016, doi: 10.1001/jama.2015.18421.

[7] D. G. Kiely, C. A. Elliot, I. Sabroe, and R. Condliffe, “Pulmonary hypertension: diagnosis and management,” *BMJ*, vol. 346, no. apr16 1, pp. f2028–f2028, Apr. 2013, doi: 10.1136/bmj.f2028.

[8] S. Greiner *et al.*, “Reliability of Noninvasive Assessment of Systolic Pulmonary Artery Pressure by Doppler Echocardiography Compared to Right Heart Catheterization: Analysis in a Large Patient Population,” *J. Am. Heart Assoc.*, vol. 3, no. 4, Aug. 2014, doi: 10.1161/JAHA.114.001103.

[9] R. L. Benza *et al.*, "Predicting Survival in Pulmonary Arterial Hypertension: Insights From the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management (REVEAL)," *Circulation*, vol. 122, no. 2, pp. 164–172, Jul. 2010, doi: 10.1161/CIRCULATIONAHA.109.898122.

[10] C. S. Johns, D. G. Kiely, and A. J. Swift, "Novel imaging techniques in pulmonary hypertension," *Curr. Opin. Cardiol.*, vol. 33, no. 6, pp. 587–593, Nov. 2018, doi: 10.1097/HCO.0000000000000559.

[11] G. E. D'Alonzo, "Survival in Patients with Primary Pulmonary Hypertension: Results from a National Prospective Registry," *Ann. Intern. Med.*, vol. 115, no. 5, p. 343, Sep. 1991, doi: 10.7326/0003-4819-115-5-343.

[12] A. Leha *et al.*, "A machine learning approach for the prediction of pulmonary hypertension," *PLOS ONE*, vol. 14, no. 10, p. e0224453, Oct. 2019, doi: 10.1371/journal.pone.0224453.

[13] P. Rajiah, "The Evolving Role of MRI in Pulmonary Hypertension Evaluation: A Noninvasive Approach from Diagnosis to Follow-up," *Radiology*, vol. 289, no. 1, pp. 69–70, Oct. 2018, doi: 10.1148/radiol.2018181080.

[14] M. E. Ari *et al.*, "Assessment of left ventricular functions and myocardial iron load with tissue Doppler and speckle tracking echocardiography and T2* MRI in patients with β-thalassemia major," *Echocardiography*, vol. 34, no. 3, pp. 383–389, Mar. 2017, doi: 10.1111/echo.13463.

[15] F. Grothues, J. C. Moon, N. G. Bellenger, G. S. Smith, H. U. Klein, and D. J. Pennell, "Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance," *Am. Heart J.*, vol. 147, no. 2, pp. 218–223, Feb. 2004, doi: 10.1016/j.ahj.2003.10.005.

[16] J. Kwon, K.-H. Kim, J. Medina-Inojosa, K.-H. Jeon, J. Park, and B.-H. Oh, "Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography," *J. Heart Lung Transplant.*, vol. 39, no. 8, pp. 805–814, Aug. 2020, doi: 10.1016/j.healun.2020.04.009.

[17] J. Peng, C. Chen, M. Zhou, X. Xie, Y. Zhou, and C.-H. Luo, "A Machine-learning Approach to Forecast Aggravation Risk in Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease with Clinical Indicators," *Sci. Rep.*, vol. 10, no. 1, p. 3118, Dec. 2020, doi: 10.1038/s41598-020-60042-1.

[18] M. T., D. Mukherji, N. Padalia, and A. Naidu, "A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)," *Int. J. Comput. Appl.*, vol. 68, no. 16, pp. 11–15, Apr. 2013, doi: 10.5120/11662-7250.

[19] M. A. Jabbar, B. L. Deekshatulu, and P. Chndra, "Alternating decision trees for early diagnosis of heart disease," in *International Conference on Circuits, Communication, Control and Computing*, Bangalore, India, Nov. 2014, pp. 322–328. doi: 10.1109/CIMCA.2014.7057816.

[20] X. Wu *et al.*, "Value of a Machine Learning Approach for Predicting Clinical Outcomes in Young Patients With Hypertension," *Hypertension*, vol. 75, no. 5, pp. 1271–1278, May 2020, doi: 10.1161/HYPERTENSIONAHA.119.13404.

[21] X.-L. Zou *et al.*, “A promising approach for screening pulmonary hypertension based on frontal chest radiographs using deep learning: A retrospective study,” *PLOS ONE*, vol. 15, no. 7, p. e0236378, Jul. 2020, doi: 10.1371/journal.pone.0236378.

[22] A. Lungu, A. J. Swift, D. Capener, D. Kiely, R. Hose, and J. M. Wild, “Diagnosis of Pulmonary Hypertension from Magnetic Resonance Imaging–Based Computational Models and Decision Tree Analysis,” *Pulm. Circ.*, vol. 6, no. 2, pp. 181–190, Jun. 2016, doi: 10.1086/686020.

[23] J. Kwon, K.-H. Kim, J. Medina-Inojosa, K.-H. Jeon, J. Park, and B.-H. Oh, “Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography,” *J. Heart Lung Transplant.*, vol. 39, no. 8, pp. 805–814, Aug. 2020, doi: 10.1016/j.healun.2020.04.009.

[24] T. J. W. Dawes *et al.*, “Machine Learning of Three-dimensional Right Ventricular Motion Enables Outcome Prediction in Pulmonary Hypertension: A Cardiac MR Imaging Study,” *Radiology*, vol. 283, no. 2, pp. 381–390, May 2017, doi: 10.1148/radiol.2016161315.