International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V71I2P210
Volume 71 Issue 2, 78-88, February 2023

Original Article

An Analytical Framework for Screening Cardiogenic Brain Abscess in Patients with Tetralogy of Fallot

Subramaniyan Mani[1], Sumit Thakar[2], N. S. Suijth[3], R. Raghunatha Sarma[4]

[1]*Hospital Management Information System, Sri Sathya Sai Institute of Higher Medical Sciences, Karnataka, India.*
[2]*Neurological Sciences, Sri Sathya Sai Institute of Higher Medical Sciences, Karnataka, India.*
[3]*Cardiothoracic and Vascular Surgery, Sri Sathya Sai Institute of Higher Medical Sciences, Karnataka, India.*
[4] *Mathematics and Computer Science, Sri Sathya Sai Institute of Higher Learning, Andhra Pradesh, India.*

[1]*Corresponding Author : subramaniyan.m@sssihms.org.in*

Received: 17 October 2022 Revised: 31 January 2023 Accepted: 11 February 2023 Published: 25 February 2023

Abstract - *Patients diagnosed with the congenital heart condition called Tetralogy of Fallot(TOF) are prone to developing cardiogenic brain abscess(CBA), the diagnosis of which is often delayed in a resource-limited setting. The current study is aimed to demonstrate the effective application of various Machine Learning(ML) techniques through appropriate data-mining strategies in the screening process for CBA. The data set for this retrospective study included clinical, echo-cardiographic and radiological variables pertaining to TOF patients with CBA. The study demonstrates four data mining tasks to highlight the importance of machine learning techniques for the screening of CBA in TOF patients. Firstly, suitable ML techniques are used to classify the TOF patients with BA correctly. Conformal prediction is then used to provide levels of reliability for individual predictions. 'SHAP' analysis is done to provide model explainability. Finally, Association Rule Mining(ARM) is utilized to draw the relationships between the top features and the outcome, thus identifying variables that expressed optimal interestingness in the study. The evaluation metrics for the ML models were better than those of the LR model, with Random forest(RF) performing the best (precision: 0.94; recall, accuracy, F-score and Area under the curve: 0.93). Conformal prediction analysis revealed an accuracy of 0.95. Association rule mining identified a combination of 'Neutrophil/Lymphocyte ratio', absence of 'branch confluence' and the 'presence of cyanosis' to have a significant 'Irule' value (1.03). Machine learning outperforms the 'BA-TOF score', a logistic regression-based score in identifying biomarkers for TOF patients. The RF algorithm demonstrated the best evaluation metrics of the various models tested. A combination of three variables can accurately and reliably aid the clinician in suspecting a CBA in male TOF patients. An ML framework designed with careful analysis to provide definitive and largely interpretable results with high confidence is proven to be an able tool in clinical decision-making.*

Keywords - *Association rule mining, Cardiogenic brain abscess, Logistic regression, Machine learning, Random forest.*

1. Introduction

Risk stratification is an important requirement in medical practice to assess a patient's health needs based on the risk tiers that the patient may belong to. Continuous assessment and developing risk scores help health practitioners predict the risk and draw out specific care plans to address the patient's health needs. Therefore it is apt to make use of Machine learning(ML) algorithms to facilitate reliable conclusions based on what it 'learns' and analyzes from large historical data sets. The main source of inspiration for this study comes from the work [1], where authors have undertaken a statistical study to identify markers for the occurrence of brain abscesses in patients diagnosed with congenital heart conditions. However, the advantages of ML over traditional statistical methods such as logistic regression(LR) include its emphasis on prediction rather than inference, its ability to handle non-linear relationships between predictors and outcomes, and the generation of more practical and reliable models [2],[3]. Though the work highlights the known risk factors, machine learning methods are still under-explored in the study. Over and above the identification of risk factors, they may be extended to give better explainability and evolve strategies in patient management. The latest work in [4] discusses the extendibility of machine learning as a strategy to develop clinical decision support. In the context of developing countries with resource-restricted settings, machine learning can further be utilized to improve risk stratification, as demonstrated in the study [5]. Though utilizing ML can be seen as a utility for assistance in decision support, its contribution and alignment as a support can become a go-to tool for clinicians for increasing diagnostic accuracy and prognosis [6],[7]. In the field of neurological sciences, ML models are increasingly being accepted as a tool that improves preoperative decision-making and measuring peri-operative outcomes [8-11]. These studies establish the use of ML models as a practical prediction tool. ML-based models in medicine are expected to translate to better outcomes at lower healthcare costs by enhancing risk stratification and clinical decision-making. However, a wider analytical framework in the clinical domain could boost the confidence of the clinical community to adopt ML tools for precision medicine.

1.1. Problem Statement

Abscess in the brain is a swelling caused by accumulated pus and dead cells. It is formed as a result of bacterial or fungal infection arising through a wound in the head or anywhere else in the body. Predominantly people with a compromised immune system, chronic illnesses, meningitis, those on immunosuppressant drugs, chronic sinusitis or middle ear infection, head injury or congenital heart diseases are prone to brain abscesses. Amongst all, congenital defects of the Heart, teeth or intestine are more susceptible compared to others. Cardiogenic brain abscess is observed and identified to be one of the major causes of mortality in children with congenital defects. Children with Tetralogy of Fallot (TOF) have four defects. They are ventricular septal defect, pulmonary stenosis, overarching aorta and right ventricular hypertrophy. This condition causes a lack of oxygen-rich blood reaching all body parts. The Intra-cardiac Right to Left shunt allows bacteria to avoid the phagocytes in the pulmonary circulation and enter the cerebral circulation. This creates an active bed for infection leading to a brain abscess. Thus delay in diagnosis of brain abscess increases the risk of mortality in these patients.

Headache, focal neurologic deficits, and perceived sensory and behavioural changes developed and observed over time, coupled with frequent chills and fever, present as symptoms to watch out for possible interrogation and diagnosis of brain abscess. In children, continuous crying, vomiting and limb spasticity, along with fontanelle swelling, could indicate a need for medical intervention. Since these symptoms are also observed for TOF in general and also other illnesses, a neurological exam along with routine blood investigations followed by a radiological examination of computed tomography(CT) or magnetic resonance imaging(MRI) exam is to be performed to confirm a brain abscess.

Regular blood investigations such as complete blood count with differential and platelet count, erythrocyte sedimentation rate, serology, C-reactive protein, and blood culture are performed to start on the required antibiotic regime. Occasionally, lumbar puncture is done to draw cerebrospinal fluid(CSF) to rule out infection. However, this is done only after considering that the procedure would not cause additional intracranial pressure, which may worsen the condition and even lead to death. Abscesses are managed medically with carefully identified antibiotics based on microorganisms isolated. Surgical intervention may be required based on the prognosis and the size and location of the abscess. The identification and early treatment are critical in the patient's interest.

The current study is an effort to build an analytical framework to identify predictive factors and carefully analyse a protocol that could be established for early identification of the occurrence of brain abscesses in TOF patients. The work considers a comprehensive set of possible characteristics that could indicate the presence of brain abscess in TOF condition.

2. Related Works, Materials and Methods

2.1. Literature, Research Gap & Novelty

In the neurosurgical field, ML has been used in diverse areas such as the preoperative planning of epilepsy surgery and glioma resection, predicting vasospasm after subarachnoid haemorrhage, diagnostic classification of lumbar disc degeneration, intra-axial tumors and prediction of post-operative satisfaction and complications [11-18]. Researcher [11] has applied to boost algorithm to preoperatively identify the risk of immediate post-operative risk in intracranial tumour surgery to assess its predictive ability in the work. They have established that the prediction model's performance was better than that of statistical analysis, wherein pathological and surgery-related parameters were found to be predictive, which were not identified by the conventional method. Work discussed in [12] carries out the pre-surgical classification of epilepsy type for localization to provide the required information. Authors have applied data mining algorithms on genetics-based features to localize symptoms of epilepsy. The results were found to perform closely to an experienced clinician in this scenario. Though the authors in [13] have attempted to arrive at a prediction of possible surgical resection in patients with glioblastoma, a clinical judgement of such a condition itself is subject to discussion because the prognosis and the treatment plan may vary with each individual surgeon.

The study's authors [19] have experimented with various feature selection methods to classify prostate cancer patients from radiological images. The work establishes the use of machine learning methods in the medical field, particularly in feature selection and classification. The work describes using ANOVA, MI, and RFE methods for feature selection. It compares the performance of various classifiers such as support vector machines(SVM), gradient boosting methods, k-Network neighbourhood, LR and Naive Bayes. The work highlights the need for risk stratification using FS methods of ML in the medical field. Work in [20] is yet another work in the pre-processing data area and feature selection for cardiovascular disease detection and classification. They have synthesized the field optimization technique for data conversion with ML to select the most appropriate feature in disease identification. In work [21], ML and deep learning methods have been reviewed to determine vaccine efficacy and outcomes.

Determining outcomes of the proposed treatment and the most predictive cause is a natural question that a medical researcher often deals with. The work by [22] discusses the importance of identifying relevant features and predicting cardiac disease. The authors have proposed a deep learning method called the extreme learning machine model to achieve the same. Breast cancer is one of the major causes of death in women across various nations. The review done in work [23] discusses the role of a comprehensive approach of considering the image-specific findings, blood work and clinical parameters applying ML techniques in detecting breast cancer. Works discussed above have extensively dealt with feature selection and analysis in the field of medicine and the role of artificial intelligence in the early detection of diseases. The works as such have been a huge inspiration for the authors to delve deeper, study and utilize techniques to draw feature importances that are clinically significant and get improved accuracy of the models.

Thus, building reliable classification algorithms for predictions, as discussed in [14],[15], boost researchers' confidence in utilizing ML techniques as a diagnostic tool in preoperative scenarios for assessing post-operative risks and lumbar disc degeneration. The work in [14] essentially has utilized an artificial neural network framework to establish the capability of advanced machine learning techniques compared to simple logistic regression models. In various clinical problems, it has also been observed that ML models outperform traditional statistical studies [25]. As observed, most of the studies referred hither-to earlier have not been studied in depth, nor have considered comprehensive factors to analyse BA in TOF condition. Moreover, no studies have employed ML tools to analyse and predict the disease condition under discussion.

Further, most of the studies in the literature reviewed have concluded with achieving good classification accuracy for the outcomes classified by the models chosen. They have not delved into extending variables obtained in the feature selection process to draw clinically relevant insights. This is referred to in the medical domain as bench-to-bedside or translational research. Hence, the current study hypothesised that ML tools would outperform traditional LR in terms of their predictive accuracy and reliability based on appropriate model selection and careful analysis, providing more practical insights. Presenting a careful analysis of a protocol for a possible early intervention is a novelty attempted in this work.

About the work [1] mentioned earlier, where a scoring mechanism based on logistic regression(LR) analysis for the prediction of the occurrence of cardiogenic brain abscess(CBA) is demonstrated, the goal of the current study is to highlight the applicability of ML in identifying markers for CBA in Tetralogy of Fallot(TOF) patients and provide an assistive clinical decision tool for effective screening.

The data set for this retrospective study was the same as that was used in the publication [1] and spanned over 15 years from 2001 to 2016. It included 30 TOF patients diagnosed with CBAs in the unit and 85 without CBAs. Patients had either not been treated or had been palliatively managed for TOF before admission at Sri Sathya Sai Institute of Higher Medical Sciences, Whitefield. Being a retrospective study that used de-identified data, the study qualified for a waiver by the Institutional review board and ethics committee.

Table 1. Descriptive statistics of the variables in the dataset

Clinical variables	Range	Average
Age	2 - 15	8
Body Mass Index	8 - 16	12.8
Body Surface Area	0.2 - 1.5	0.76
Cyanotic spells	Yes / No	No
Persistent cyanosis	present/absent	present
Dyspnea on exertion	present/absent	present
Age-adjusted heart rate percentile	20 - 99	60
SpO2	66 - 98	77
Echocardiographic variables		
Pulmonary artery index	6 - 51	19.5
Pulmonary stenosis type	valvular / infundibular/ both	both
Branch confluence	present/absent	present
Pulmonary valve gradient	50 - 130	80
VSD type	membranous / peri-membranous / muscular	
Laboratory variables		
PCV	31 - 82	56
Total Count	4200 - 21700	9955
Neutrophil/ lymphocyte count	0.1 - 10	2.02
Platelet count	1 - 5.3	2.32
ESR	1 - 82	5.32
Direct Bilirubin	0.02 - 0.7	0.14
Indirect Bilirubin	0.08 - 1.72	0.78
Prothrombin time	12 - 60	19.7
Activated partial thromboplastin time	14 - 70	36.7
Bleeding time	1.5 - 4	2.5
Clotting time	3.5 - 11.5	5.9
Radiographic variable		
Pulmonary oligemia on CXR	present/absent	absent

(Abbreviations : (SpO2= oxygen saturation, VSD= ventricular septal defect, CXR = chest X-ray)

2.2. Predictors included in the Study

The variables included the various clinical, laboratory, ECHO cardiography and chest radiograph data. They are listed in Table-1. These variables have been chosen based on the symptoms identified for the condition and based on the treatment protocol for both brain abscess and TOF conditions in the literature and clinical practice. Since the risk of BA in an untreated TOF is expected to be high, variables reflecting the manifestation of a severe TOF have been considered. Data were collected from patient charts and from the electronic medical records available in the hospital. Table-1 also contains the descriptive statistics detailing the range of values and aggregate for the variables in the data set.

2.3. Data Analysis

Obtaining data in the medical domain is often challenging and more so labelled data in a clinical problem spanning across specialities. The existence of a structured electronic medical record eases the data collection to some extent and could fast-track the data preparation step. In the medical domain, classification tasks often suffer from the inherent data imbalance problem due to the low availability of positive cases. Re-sampling techniques such as oversampling and under-sampling are widely used to handle data imbalance, while with the caution of bias that this may introduce. Synthetic Minority Over Sampling(SMOTE) technique is a widely used technique with various adaptations for handling the imbalance problem [26],[27]. SMOTE generates artificial minority class data samples by interpolation between selected nearest minority neighbours. Ensemble classification techniques perform relatively well in an imbalanced scenario, though originally not designed to handle imbalance problems[28]. The inherent capability of ensemble techniques of handling bias-variance trade-offs favour their usage. Therefore, a simple SMOTE coupled with ensemble classification techniques has been experimented with in the current work. To handle class imbalance (approximately 74% to 26%) in the patient data, the experiments illustrated the effectiveness of SMOTE in handling class imbalance and obtaining a good performance.

2.3.1. Classification Task

The exclusivity of the data set has resulted in limited options to deal with the classification tasks. As the need was to build a risk stratification model, obtaining the top predictive features with good accuracy was imperative. Several ML models are available and can be trained on the data. Owing to the scarcity of data, a cross-validation approach was considered. The purpose of cross–validation is to validate the ability of the ML models on unseen data to predict the output. It is also used to mitigate problems of over-fitting and selection bias. It also helps one understand the model's capability to perform on a new unseen dataset.

Support vector machines (SVM) is one of the ML methods with a strong statistical learning theory to its credit available to process complex data with good accuracy. It is typically used for cases where data sets are small, and it is difficult to estimate the underlying distributions. It works well and does not suffer much due to the high dimensionality of the data. It finds a hyperplane that separates the data into two classes while maximizing the margin of decision boundary separating the two classes with minimal classification errors.

Ensemble techniques try to improve classification accuracy by combining and aggregating the predictions of the base classifiers. Thus, these techniques are also described as classifier combination methods. Since an ensemble of classifiers tends to decrease generalization error, it may result in better performance than single classifiers, specifically when there is a low correlation among the base classifiers. Random Forest (RF), a variant in the bagging technique, is one of the most used algorithms because of its simplicity. Subsets of the original data set are considered with replacement. All subsets are maintained the same size, and there is a high probability that different data points are provided to the models. The final output is typically based on the majority votes of predictions. The algorithm of Random Forest is presented in Algorithm-1.

Algorithm-1: Random Forest

Precondition: Training set $T = (x_1, y_1), \ldots, (x_n, y_n)$, features K and the no of trees in the Forest B

function RANDOMFOREST(T,K) :
    $M \leftarrow \emptyset$
    for $i \in 1, \ldots, B$ **do**
        $T^{(i)} \leftarrow$ A bootstrap sample from T
        $h_i \leftarrow LEARN\text{RANDOMIZEDTREE}\ (T^{(i)}, K)$
        $M \leftarrow M \cup \{h_i\}$
    end for
    return *M*
end function

function LEARNRANDOMIZEDTREE(T,K) :
    At every node :
        $k \leftarrow$ a subset of K
        Split on the best feature in *k*
    return The Learned Tree
end function

2.3.2. Confidence Analysis

In a high-risk setting such as a medical diagnosis that demands uncertainty quantification, a method for creating confidence intervals/sets are found to be helpful since they overcome distributional/model assumptions even with a small data cohort. Conformal prediction(CP) is a technique to quantify uncertainty by creating statistical sets/intervals for predictions produced by the underlying ML model [29]. The CP technique estimates a set of classes for the classification problem for a given input. The prediction sets obtained guarantee to contain the true classification class with high probability[30][31]. The algorithm for Inductive conformal prediction is presented in Algorithm-2.

Algorithm-2*: *Inductive Conformal Prediction

***Input**:* Training data $(x_i, y_i)_{i\epsilon I_{train}}$, calibration data $(x_i, y_i)_{i\epsilon I_{calib}}$, Y the set of possible labels, significance level α, classifier C, non-conformity score r, new point x_{new}

Output : $C_{conf}(x_{new})$, prediction region at x_{new}

Fit the classifier on training data C on $(x_i, y_i)_{i\epsilon I_{train}}$
Compute the non-conformity score on calibration data $(r(x_i, y_i))_{i\epsilon I_{calib}}$

for $y \in Y$ **do**
a) Let $\tilde{y}_{new}$ the probability predicted by C for x_{new} to belong to class y;
b) Compute $r(x_{new}\ , y) = 1 - \tilde{y}_{new}$ and the associated p-value $p_y(x_{new})$
end
return $C_{conf}(x_{new}) = \{y \in Y \mid p_y(x_{new}) \geq \alpha$

Thus, classification results tabulated with the conformity measures provide high confidence to proceed further in the medical domain.

2.3.3. Model Explainability

Explainable AI, the set of ML techniques that offer interpretability, is much sought after these days due to the trust demanded in the medical domain. It also gives better scope for the rapid adoption of AI to make implementable decisions. For example, identifying the features' importance and developing the ML models provide information about the features that impact the outcome. For a random forest technique, one can identify and present how each feature decreases the impurity of the split on average. The average of this calculation over the trees in the forest gives the score to measure the feature's importance.

While a variety of tools offer visualization and descriptions for AI models, Shapely Additive Explanations (SHAP) provide good explainability with its global and local interpretability capability[32]. SHAP values provide the contribution of every feature towards the expected model prediction corresponding to every row of the data set [32]. The SHAP implementation draws its inspiration from cooperative game theory, where a coalition of players, being the units of decision-making, look towards optimizing their payouts/results in the game. In the context of ML, each feature for a given sample could be considered a participant in a cooperative game. The where-in prediction could be interpreted as the payout. If Z(i) is the importance of a feature, the effect on the predictions of including Z(i) to the subsets of features not containing Z(i) is first estimated, and all contributions are aggregated to derive the marginal contribution of the feature[33]. However, here, SHAP does not retrain the model for each of these subsets. Instead, the average value of the feature is substituted, and predictions are generated for the left-out feature. SHAP utilizes the average calculated over marginal contributions across all permutations. Global interpretability is inferenced by the collective SHAP value obtained for each predictor's positive or negative contribution to the target variable. Individual SHAP values contribute to local interpretability. The features with their SHAP values are presented in descending order highlighting the feature's importance to the target variable.
In the current work, SHAP was used to provide the model explainability by identifying factors that impact the outcome positively or negatively.

2.3.4. Actionable Knowledge with Combined Pattern Matching

Association rule mining(ARM) is a paradigm in ML used to identify interesting associations and relationships that exists in large sets of data items.[39,42] Typical ARM produces large sets of rules, which may make it time-consuming and challenging to interpret and understand for users. In the medical domain, ARM could be utilized to draw the relationships between the top features and the outcome. The methodology of considering top features reduces the computation complexity. Apart from the top features that could be combined, the concept of combined patterns to extract actionable knowledge from the generated set of rules was explored in the current work. Combined patterns consist of association rules and rule pairs that are combined[34]. They can be utilized to derive actionable knowledge that is interesting and implementable. The concepts described for arriving at combined patterns, association rules and their pairs are briefly introduced here. The combined association rule involves generating multiple heterogeneous item sets from different datasets. In the current context, clinical, echo and radiological variables were considered the first dataset, and the demographic information was considered the second data set. The association rules were combined with looking for its '*interestingness'* to the outcome target variable under study.

Combined Association Rule- CAR:

Let there be n datasets K_i , for $i = (1..n)$. Let M_i indicate the set of all items in datasets K_i and ∀i = j, $M_i \cap M_j = \emptyset$. A combined association rule R can be described as
$C_1 \wedge C_2 \wedge \ldots \wedge C_k \rightarrow L$, where $C_i \subseteq M_i\ (i = 1..n)$ is an itemset in the data set K_i , $L \neq \emptyset$ is a target item or class, and ∃ i,j, i≠ j, $C_i \neq \emptyset$, $C_i \neq \emptyset$
Rule pairs are generated by combining similar but contrasting rules that are already obtained as follows:

Interestingness of Combined Association Rules:

Apart from regularly used measures such as support, confidence and lift, the interestingness of combined association rules is being introduced by defining two lifts to measure the interestingness.

Table 2. Result Metrics for Classification task

Model	Precision	Recall	Accuracy	F Score	AUC
Linear SVM	0.89	0.9	0.89	0.89	0.89
RBF SVM	0.92	0.91	0.91	0.91	0.91
Poly SVM	0.88	0.93	0.9	0.9	0.9
XGB	0.93	0.92	0.92	0.92	0.92
RF	0.94	0.93	0.93	0.93	0.93

Lift is the ratio of confidence of the rule & expected confidence of the rule. It measures the performance of an association rule to predict or classify data items with more impact or the ones having an enhanced response compared to the entire population[34].

$$Lift_X(X \wedge Y \rightarrow \mathrm{L}) = \frac{Conf(X \wedge Y \rightarrow L)}{Conf(Y \rightarrow L)} = \frac{Lift(X \wedge Y \rightarrow L)}{Lift(Y \rightarrow L)} \quad (1)$$

$$Lift_Y(X \wedge Y \rightarrow \mathrm{L}) = \frac{Conf(X \wedge Y \rightarrow L)}{Conf(X \rightarrow L)} = \frac{Lift(X \wedge Y \rightarrow L)}{Lift(X \rightarrow L)} \quad (2)$$

$Lift_X(X \wedge Y \rightarrow \mathrm{L})$ gives the lift of X with Y as a precondition, which shows how much X contributes to the rule. Likewise $Lift_Y(X \wedge Y \rightarrow \mathrm{L})$ gives the contribution of Y in the rule.

The interestingness of the combined association rules is defined using the two new lifts defined above.

$$I_{rule}(X \wedge Y \rightarrow \mathrm{L}) = \frac{Lift_X(X \wedge Y \rightarrow L)}{Lift(X \rightarrow L)} \quad (3)$$

$$I_{rule}(X \wedge Y \rightarrow \mathrm{L}) = \frac{Lift(X \wedge Y \rightarrow L)}{Lift(X \rightarrow L) Lift(Y \rightarrow L)} = \frac{Lift_Y(X \wedge Y \rightarrow L)}{Lift(Y \rightarrow L)} \quad (4)$$

Here I_{rule} indicates if the contribution of A given B to the occurrence of L increases or not. Therefore, from "$I_{rule} < 1$" it can be concluded that A∧B → L is less interesting than A → L and B → L. The I_{rule} values are in the range [0,+∞). When $I_{rule} > 1$, the higher I_{rule} indicates the higher interestingness of the rule. Here I_{rule} measures the quantitatively the interestingness of a rule. The measure indicates how unexpected may be a combined rule compared to simple rules produced in basic association analysis. CAR was used to generate combined patterns and identify the rules with $I_{rule} > 1$ that are useful as predictive factors for the outcome.

In summary, a four-step approach was adopted to achieve the study's objective.

1. Classification Task- Identification of suitable classifier
2. Confidence Analysis- A paradigm such as an inductive conformal prediction framework for providing confidence with regard to the chosen classifier
3. Model Interpretability- SHAP analysis highlights the predictive features with local and global interpretability regarding the outcome.
4. Actionable Knowledge Mining- Combined associative rule mining to identify *interesting* patterns of the predictive features to arrive at a clinical decision.

3. Results and Discussion

3.1. Results

The classification task experimented with SVMs with linear, polynomial and radial basis function(RBF) kernels. Following this, ensemble learners such as random forest and extreme gradient boosting (XGB) were attempted. The data was trained and fitted with 3,5, 7, and 10-fold cross-validation. Cross-validation with 5 folds gave an improved outcome. The measures are tabulated in Table-2. The random forest gave the best results with an F-score of 0.93 and an AUC of 0.93. Then conformal predictors as a wrapper for the classification were applied to generate the sets of classes to quantify the prediction obtained with reasonable accuracy. The following results were obtained by applying conformal prediction to the data set:

The average false p-value is: 0.0219
The accuracy of prediction: 0.9523
The test error rate is: 0.0476

The conformal calibration curve corresponding to the confidence analysis is presented in Figure-1. The graph has the fraction of error on the Y axis and the significance level on the X axis. It can be observed in the figure that the curve of the conformal predictor achieved by joining the orange dots is closer to the perfectly calibrated curve indicated by the straight blue line. Hence it is well-calibrated. Having achieved good confidence in the classification task, model explainability and data visualization were obtained by SHAP plots. The SHAP summary plot is presented in figure-2.

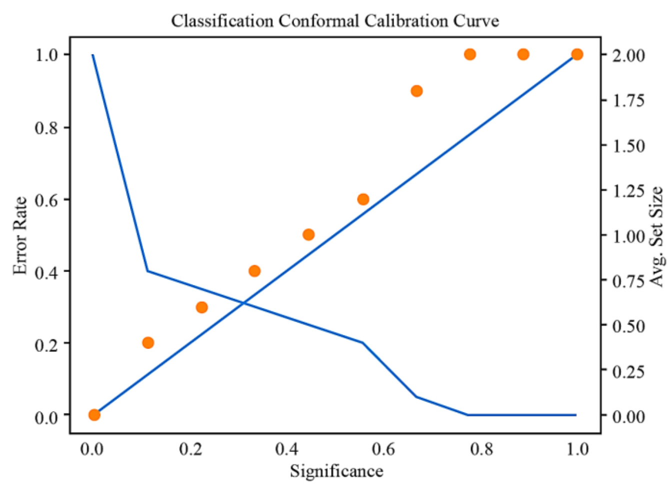

Fig. 1 Conformal prediction calibration curve

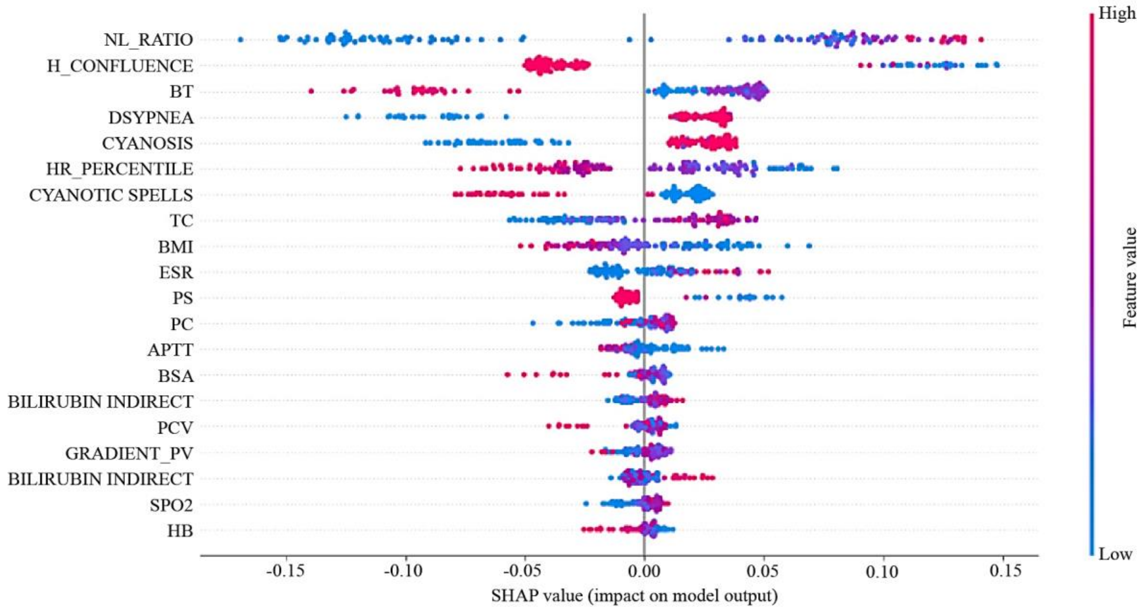

Fig. 2 SHAP Analysis

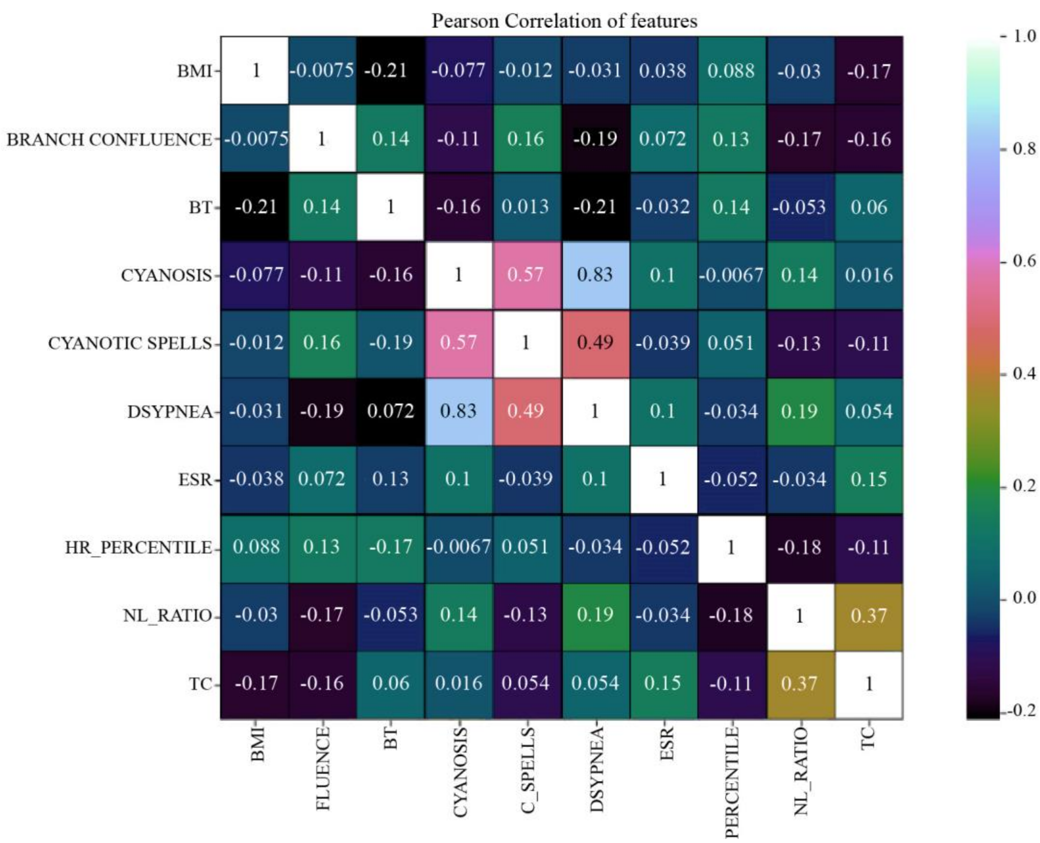

Fig. 3 Pearson correlation of top 10 features

Table 3. Result of Combined pattern matching

Rules	Gender	Outcome	Support	Confidence	Lift	Lift1	Lift2	Irule
('0','C1', 'NL4')	1	1	0.140351	0.8	2.941	1.104	2.747	1.031

Feature importance: Variables are displayed with their ranks in the descending order

Impact: The horizontal axis shows if the variable is positively or negatively correlated with the outcome

Original value: Colour indicates if the value of the variable is higher (in red) or lower (in blue) for that observation

Correlation: A higher level of the “NL_Ratio” has a high and positive impact on the outcome, the occurrence of brain abscess. On the X-axis, the red colour indicates the “High” and the “positive” impact. Likewise, the “Branch confluence” is negatively correlated with the outcome. That is, the absence of branch confluence is correlated with the occurrence of brain abscess. Further, the pearson correlation signifying the interplay between the features also was attempted and is presented in the figure-3. Only cyanosis and cyanotic spell have a marginally significant positive correlation. However, not every cyanotic patient develops a cyanotic spell.

The next task was to generate combined patterns. The top 14 features out of 28, suggested by the feature importance function and the SHAP plot, were identified for associative rule mining. In order to clearly identify the features most indicative of risk in the patient cohort, the idea of combined patterns to extract useful and actionable information from existing rules was attempted. The top 14 features were split into First-7 & Second -7 data sets and given as input for 2 experiments. The First-7 are NL ratio, branch confluence, bleeding time(BT), dyspnoea, cyanosis, heart rate percentile and cyanotic spells. The patient cohort was partitioned into two groups, male and female and the two groups were mined separately. Combined rules were generated, and metrics discussed earlier, such as Lift and Interestingness, were used to determine the feature influencing the outcome. Rules with change in outcome(presence of brain abscess) were obtained by identifying rules with Irule greater than or equal to 1. The First-7 yielded interesting results, while the Second-7 did not have any rule, with Irule greater than 1. The result is presented in table Table-3.

The value '0' correspond to the absence of branch confluence, 'C1' for the presence of cyanosis, 'NL4' for NL ratio values greater than or equal to 4, 'BT3' for BT values between 3 and 4, 'CS0' for the absence of cyanotic spells. It can be observed that the absence of branch confluence with elevated NL ratio is indicative of brain abscess for the cyanotic male patients with TOF condition are suggestive of high risk for brain abscess.

3.2. Discussion

The LR-based ‘BA-TOF’ score was an effort to identify biomarkers that could potentially aid primary-care physicians in suspecting CBAs and stepping up appropriate referrals for brain CT scans[1]. Using three variables (N/L ratio, ESR and heart-rate percentile), the score generated by the model can correctly predict a CBA in more than 80% of the TOF patients. Building on this further, the results of the current study indicate that ML tools (SVM, XGB and RF) outperform traditional LR on various evaluation metrics (precision, recall, accuracy, F-score and AUC) (Table-2).

ML-based models would, hence, be more accurate and reliable than the ‘BA-TOF’ score in identifying TOF patients likely to be harbouring a CBA. Amongst the various ML tools used in this study, RF demonstrated the best evaluation metrics. The robustness of this ML tool has been previously validated in a systematic review of 48 studies that use varied ML approaches and methods for disease prediction. It was found that while SVM was applied most frequently, the RF algorithms demonstrated the best accuracy[35]. This has been ascribed to various advantages of RF over other ML tools, such as resilience to noise and its ability to deal with multimodal data and evaluate biomarkers [36].

In order to adopt ML predictive models in clinical practice, it is often reiterated that the models should cater to individual patients and not just the population as a whole. Confidence analysis using conformal predictions is a promising approach that provides certainty levels for individual predictions and is increasingly being used to augment the utility of ML tools [37]. In the current study, CP analysis revealed an accuracy rate of 95% and an error rate of 0.04, indicating that the RF model can be used with a very high confidence level to predict CBA in a given TOF patient.

SHAP breaks down the mechanics of an ML model and renders it easily understandable. In the current study, N/L ratio, dyspnoea and cyanosis had the highest SHAP values and strong positive correlations with the occurrence of a CBA. At the same time, the presence of a branch confluence of the pulmonary arteries and bleeding time demonstrated strong negative correlations (Figure-2). Interestingly, the three variables of the ‘BA-TOF’ score (N/L ratio, heart rate percentile and ESR) figured high up on the SHAP plot (Figure-2), though only the N/L ratio made it to the top three rankings.

After obtaining the results of the SHAP analysis, the significant relationships in our data set were evaluated using association analysis, an important aid in decision-making. Thus, although the N/L ratio, dyspnoea and cyanosis had the highest SHAP scores, the clinically relevant variables with the strongest relationship were: N/L ratio of more than 4, absence of branch confluence and the presence of cyanosis. While the elevation of the N/L ratio, like other inflammatory markers like ESR and C-reactive protein, is known to occur in several inflammatory conditions, the other two variables indicate a worse cardiac ‘outlet obstruction’ and hence a more severe TOF pathology[1],[38]. It should be cautioned that the ML results from the association analysis hold good only for the male gender, unlike the 'BA-TOF' score that applied to either gender. Larger sample sizes may, however, yield significant results for both genders.

Translating our ML results to a real-life scenario, if a physician in a primary health clinic were to come across a cyanotic male TOF patient with non-specific headache and fever, he/she would just require his baseline echocardiography records and blood count report to rule in or rule out a CBA. If the patient's echocardiography records reveal the absence of a branch confluence, and his blood test demonstrates an N/L ratio of more than 4, the physician would be more than 90 % certain that the patient is harbouring a CBA. The patient could then be referred urgently to a higher centre for CT confirmation of the diagnosis and early institution of treatment.

3.3. Study Limitations

This retrospective study has the inherent limitation of confounding factors that are not recognised or apparent yet. The study's strength could be explored further by prospectively validating the ML models on larger samples and across different centres. ML predictive analysis was utilized in the current study to identify biomarkers that would indicate the concurrent presence of a CBA in a TOF patient. It would also be worthwhile to prospectively investigate if ML could be used to predict the future occurrence of a brain abscess in a given TOF patient based on his or her baseline investigations. It would serve as an additional pointer for the patient to seek corrective treatment for TOF as soon as possible since even a partially managed TOF patient is still at risk of developing a CBA.

4. Conclusion

Machine learning techniques have been applied to identify biomarkers and cardiogenic screen brain abscesses in the Tetralogy of Fallot patients. They outperform the traditional logistic regression-based 'BA-TOF score' in identifying the biomarkers. The Random forest algorithm demonstrated the best evaluation metrics of the various models tested. A wrapper designed with conformal prediction to rule out generalization errors boosts the confidence of the classification results obtained. SHAP analysis highlights the predictive features with local and global interpretability about the outcome. The analysis clearly presents the interplay and dependencies of the variables in this clinical scenario. Finally, the combined associative rule matching aided the clinical decision-making in arriving at a simple checklist and guidance in screening brain abscesses amongst TOF patients considered in the study. A combination of three variables can accurately and reliably aid the clinician in suspecting a CBA in male TOF patients. Thus an ML framework designed with careful analysis in providing definitive and largely interpretable results with high confidence is proved to be an able tool in clinical decision making.

Conflicts of Interest

The authors confirm that they have read the Journal's directives and understand the stand on ethical publication. The authors confirm that this report is consistent with the guidelines of the World Medical Association Declaration of Helsinki. The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors have not received any support in the form of a grant from any agency for this study or article preparation. This paper has not been published before.

Acknowledgments

The manuscript has been prepared as per institutional guidelines. The Institutional Review Board and Ethics Committee have approved the study. This work is an outcome of research collaboration between the Department of Mathematics and Computer Science, Sri Sathya Sai Institute of Higher Learning(SSSIHL,www.sssihl.edu.in)and Sri Sathya Sai Institute of Higher Medical Sciences, Whitefield(SSSIHMS-WFD,https://sssihms.org/)

References

[1] Vidyasagar Kanneganti et al., "Clinical and Laboratory Markers of Brain Abscess in Tetralogy of Fallot ('BA-TOF' Score): Results of a Case-Control Study and Implications for Community Surveillance," *Journal of Neurosciences in Rural Practice,* vol. 12, no. 2, pp. m302-307, 2021. *Crossref,* https://doi.org/10.1055/s-0041-1722819

[2] Tim Smolea et al., "A Machine Learning-based Risk Stratification Model for Ventricular Tachycardia and Heart Failure in Hypertrophic Cardiomyopathy," *Computers in Biology and Medicine,* vol. 135, p. 104648, 2021. *Crossref,* https://doi.org/10.1016/j.compbiomed.2021.104648

[3] Dan Li et al., "Machine Learning-aided Risk Stratification System for the Prediction of Coronary Artery Disease," *International Journal of Cardiology,* vol. 326, pp. 30-34, 2021. *Crossref,* https://doi.org/10.1016/j.ijcard.2020.09.070

[4] Damien K. Ming et al., "Applied Machine Learning for the Risk-stratification and Clinical Decision Support of Hospitalised Patients with Dengue in Vietnam," *PLOS Digit Health*, vol. 1, no. 1, 2022. *Crossref,* https://doi.org/10.1371/journal.pdig.0000005

[5] Justin J Boutilier et al., "Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis," *Journal of Medical Internet Research,* vol. 23, no. 1, 2021. *Crossref,* https://doi.org/10.2196/20123

[6] Brett K. Beaulieu-Jones et al., "Machine Learning for Patient Risk Stratification: Standing on, or Looking over, The Shoulders of Clinicians?," *Digital Medicine,* vol. 4, no. 62, 2021. *Crossref,* https://doi.org/10.1038/s41746-021-00426-3

[7] Ziad Obermeyer, and Ezekiel J. Emanuel, "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine," *The New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016. *Crossref,* https://doi.org/10.1056%2FNEJMp1606181

[8] Joeky T. Senders et al., "Machine Learning and Neurosurgical Outcome Prediction: A Systematic Review," *World Neurosurgery,* vol. 109, pp. 476–486, 2018. *Crossref,* https://doi.org/10.1016/j.wneu.2017.09.149

[9] Victor E. Staartjes et al., "Deep Learning-based Preoperative Predictive Analytics for Patient-reported Outcomes following Lumbar Discectomy: Feasibility of Center-specific Modeling," *The Spine Journal,* vol. 19, no. 5, pp. 853–861, 2019. *Crossref,* https://doi.org/10.1016/j.spinee.2018.11.009

[10] Victor E. Staartjes et al., "Neural Network-based Identification of Patients at High Risk for Intraoperative Cerebrospinal Fluid Leaks in Endoscopic Pituitary Surgery," *Journal of Neurosurgery,* pp. 1–7, 2019. *Crossref,* https://doi.org/10.3171/2019.4.jns19477

[11] Christiaan H.B. Van Niftrik et al., "Machine Learning Algorithm Identifies Patients at High Risk for Early Complications after Intracranial Tumor Surgery: Registry-Based Cohort Study," *Neurosurgery,* vol. 85, no. 4, pp. E756-E764, 2019. *Crossref,* https://doi.org/10.1093/neuros/nyz145

[12] Yohannes Kassahun et al., "Automatic Classification of Epilepsy Types Using Ontology-based and Genetics-based Machine Learning," *Artificial Intelligence in Medicine,* vol. 61, no. 2, pp. 79–88, 2014. *Crossref,* https://doi.org/10.1016/j.artmed.2014.03.001

[13] Adam P. Marcus et al., "Improved Prediction of Surgical Resectability in Patients with Glioblastoma using an Artificial Neural Network," *Scientific Reports,* vol. 10, no. 5143, 2020. *Crossref,* https://doi.org/10.1038/s41598-020-62160-2

[14] Travis M. Dumont, Anand I. Rughani, and Bruce I. Tranmer, "Prediction of Symptomatic Cerebral Vasospasm after Aneurysmal Subarachnoid Hemorrhage with an Artificial Neural Network: Feasibility and Comparison with Logistic Regression Models," *World Neurosurgery,* vol. 75, no. 1, pp. 57–63, 2011. *Crossref,* https://doi.org/10.1016/j.wneu.2010.07.007

[15] C.W. Pfirrmann et al., "Magnetic Resonance Classification of Lumbar Intervertebral Disc Degeneration," *Spine,* vol. 26, no. 17, pp. 1873–1878, 2001. *Crossref,* https://doi.org/10.1097/00007632-200109010-00011

[16] K. Yamashita et al., "Performance Evaluation of Radiologists with Artificial Neural Network for Differential Diagnosis of Intra-axial Cerebral Tumors on MR Images," *AJNR American Journal of Neuroradiology,* vol. 29, no. 6, pp. 1153–1158, 2008. *Crossref,* https://doi.org/10.3174/ajnr.a1037

[17] Parisa Azimi et al., "Use of Artificial Neural Networks to Predict Surgical Satisfaction in Patients with Lumbar Spinal Canal Stenosis: Clinical Article," *Journal of Neurosurgery, Spine*, vol. 20, no. 3, pp. 300–305, 2014. *Crossref,* https://doi.org/10.3171/2013.12.spine13674

[18] Loftus CM, Osenbach RK, and Biller J, Diagnosis and Management of Brain Abscess. In: Wilkins RH, Rengachary SS, eds. *Neurosurgery*. New York: McGraw-Hill; 1996:3285–3298

[19] M. N. Rajesh, B. S. Chandrasekar, and S. Shivakumar Swamy, "Prostate Cancer Detection using Radiomics-based Feature Analysis with ML Algorithms and MR Images," *International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 42-58, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I12P206

[20] S. Sivasubramaniam, and S. P. Balamurugan, "Nature Inspired Optimization with Hybrid Machine Learning Model for Cardiovascular Disease Detection and Classification," *International Journal of Engineering Trends and Technology,* vol. 70, no. 12, pp. 127-137, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I12P214

[21] Abdullah M. Baqasah, "A Survey of Machine Learning Methods to Detect COVID-19 Severity, Mortality, and Vaccines Efficacy," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 28-46, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I11P204

[22] Nandakumar Pandiyan, and Subhashini Narayan, "Prediction of Cardiac Disease using Kernel Extreme Learning Machine Model," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 364-377, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I11P238

[23] Sudha Prathyusha Jakkaladiki, and Filip maly, "Potential Role of Artificial Intelligence in Breast Cancer Detection - A Review," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 130-139, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I7P214

[24] Atul Kumar Ramotra, and Vibhakar Mansotra, "Feature Raking and Stacked Sparse Autoencoder based Framework for the Prediction of Breast Cancer," *International Journal of Engineering Trends and Technology*, vol. 70, no. 5, pp. 103-110, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I5P213

[25] Xuan Songa et al., "Comparison of Machine Learning and Logistic Regression Models in Predicting Acute Kidney Injury: A Systematic Review and Meta-analysis," *International Journal of Medical Informatics*, vol. 151, p. 104484, 2021. *Crossref*, https://doi.org/10.1016/j.ijmedinf.2021.104484

[26] Nitesh V. Chawla et al., "SMOTE : Synthetic Minority Oversampling Technique," *Journal of Artificial intelligence research,* vol. 16, 2002. *Crossref,* https://doi.org/10.1613/jair.953

[27] Koichi Fujiwara et al., "Over- and Under-sampling Approach for Extremely Imbalanced and Small Minority Data Problem in Health Record Analysis," *Frontiers Public Health*, vol. 8, 2020. *Crossref*, https://doi.org/10.3389/fpubh.2020.00178

[28] Mikel Galar et al., "A Review on Ensembles for the Class Imbalance Problem: Bagging, Boosting-, and Hybrid-based Approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Review)*, vol. 42, no. 4, pp. 463–484, 2012. *Crossref*, https://doi.org/10.1109/TSMCC.2011.2161285

[29] VN Balasubramanian et al., "Support Vector Machine Based Conformal Predictors for Risk of Complications following a Coronary Drug Eluting Stent Procedure," *36th Annual Computers in Cardiology*, pp. 5−8, 2009.

[30] Anastasios N. Angelopoulos, and Stephen Bates, "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification," *Machine Learning*, 2022. *Crossref*, https://doi.org/10.48550/arXiv.2107.07511

[31] Glenn Shafer, and Vladimir Vovk, "A Tutorial on Conformal Prediction," *Journal of Machine Learning Research*, vol. 9, pp. 371-421, 2008.

[32] Scott M. Lundberg, and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *31st Conference on Neural Information Processing Systems*, pp. 4768-4777, 2017.

[33] AnkitChoudhary, A Unique Method for Machine Learning Interpretability: Game Theory & Shapley Values, Analytics Vidya, 2019. [Online]. Available: https://www.analyticsvidhya.com/blog/2019/11/shapley-value-machine-learning-interpretability-game-theory/

[34] Yanchang Zhao et al., "Combined Pattern Mining: From Learned Rules to Actionable Knowledge," *AI 2008: Adavances in Artificial Intelligence*, pp. 393-403, 2008. *Crossref*, https://doi.org/10.1007/978-3-540-89378-3_40

[35] Shahadat Uddin et al., "Comparing Different Supervised Machine Learning Algorithms for Disease Prediction," *BMC Medical Informatics and Decision Making*, vol. 19, 2019. *Crossref*, https://doi.org/10.1186/s12911-019-1004-8

[36] Minseok Song et al., "Diagnostic Classification and Biomarker Identification of Alzheimer's Disease with Random Forest Algorithm," *Brain Sciences*, vol. 11, no. 4, p. 453, 2021. *Crossref*, https://doi.org/10.3390/brainsci11040453

[37] Janette Vazquez, and Julio C. Facelli, "Conformal Prediction in Clinical Medical Sciences," *Journal of Healthcare Informatics Research*, vol. 6, pp. 241-252, 2022. *Crossref*, https://doi.org/10.1007/s41666-021-00113-8

[38] Fauzia Imtiaz et al., "Neutrophil Lymphocyte Ratio as a Measure of Systemic Inflammation in Prevalent Chronic Diseases in Asian Population," *International Archives of Medicine*, vol. 5, no. 1, 2012. *Crossref*, https://doi.org/10.1186/1755-7682-5-2

[39] Peng ping, and Qiang, "Discretion Conserving Mining of Association Rules," *SSRG International Journal of Mobile Computing and Application*, vol. 4, no. 3, pp. 1-5, 2017. *Crossref*, https://doi.org/10.14445/23939141/IJMCA-V4I6P101

[40] Abhinav Shashank, The Importance of Risk Stratification in Population Health Management, Innovaccer Inc., 2017. [Online]. Available: https://hitconsultant.net/2017/03/13/risk-stratification-population-health-management/

[41] Christo I. Tchervenkov et al., "The Improvement of Care for Paediatric and Congenital Cardiac Disease across the World: A Challenge for the World Society for Pediatric and Congenital Heart Surgery," *Cardiology in the Young*, vol. 18, no. 2, pp. 63–69, 2008. *Crossref*, https://doi.org/10.1017/s1047951108002801

[42] Kareem Mohamed, and Ümmü Altan Bayraktar, "Artificial Intelligence in Public Relations and Association Rule Mining as a Decision Support Tool," *SSRG International Journal of Humanities and Social Science*, vol. 9, no. 3, pp. 23-32, 2022. *Crossref*, https://doi.org/10.14445/23942703/IJHSS-V9I3P105

[43] Suhas Udayakumaran, Chiazor U. Onyia, and R. Krishna Kumar, "Forgotten? Not yet. Cardiogenic Brain Abscess in Children: A Case Series-Based Review," *World Neurosurgery*, vol. 107, pp. 124–129, 2017. *Crossref*, https://doi.org/10.1016/j.wneu.2017.07.144