Predictions of Heart Diseases using An Adapted Hybrid Intelligent Framework

Sumit Kumar Soni
Research Scholar
Madhyanchal Professional University, Bhopal

Kalpana Rai, PhD
Associate Professor
Madhyanchal Professional University, Bhopal

Harsh Mathur, PhD
Associate Professor
Rabindra Nath Tagore University, Bhopal

ABSTRACT
In the last few decades, heart disease has become much more common in people of all ages, so early detection became important. There are some things that make it harder to find heart disease, like diabetes, high blood pressure, an irregular heart rate, high cholesterol, and so on. To treat heart patients effectively, it is important to be able to properly diagnose heart disease before a heart attack happens. Machine learning-based noninvasive technology can swiftly and effectively identify heart disease patients. A machine-learning-based cardiovascular disease prediction system developed using heart disease datasets in the proposed research. Cross-validation used to evaluate machine learning, feature selection, and classifiers for accuracy and specificity. Here rapidly distinguish heart patients from healthy persons using technology. Receiver optimistic curves and area under the curves for each classifier were analyzed. Classifiers, feature selection algorithms, preprocessing methods, validation procedures, and performance measurements are covered in this study. A subset and the full set of features were used to test the suggested system's performance. Recall, F1 score, and false positive rate are compared. Decreases in the number of features utilized to classify affect accuracy and runtime. An expected machine-learning-based decision support system would help clinicians diagnose heart disease more accurately.

Keywords
Heart disease, Patients, Machine learning, Decision support system, Clinicians diagnose.

1. INTRODUCTION
A disease is an abnormal medical condition that occurs in the human body. It has a detrimental impact on the functioning status of the human bodily organism . It is commonly connected with little symptoms of sickness in the patient's body[1]. According to the World Health Organization (WHO), an estimated 17 million people have died each year from cardiovascular illness, primarily heart attacks and strokes, in the previous 15 years. Heart disease is a major cause of death, so many studies have been done on it. The chances of a treatment working better are higher when heart prediction is done automatically[2]. This is a big job. Cardiovascular disease signs and symptoms are used to judge how bad a patient's situation is. People who are already more likely to get cardiovascular disease—for example, those who smoke, are overweight, or have high blood pressure are more likely to get it.

Cardiovascular diseases (CVDs) encompass a range of conditions affecting the heart and blood arteries, such as coronary heart disease, cerebrovascular illness, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis, and pulmonary embolism. In 2016, the number of fatalities caused by cardiovascular diseases (CVDs) was predicted to be around 17.9 million, accounting for approximately 31% of the total worldwide mortality rate. Among the fatalities analyzed, a significant proportion of 85% were attributed to cardiovascular events such as heart attack and stroke[3].

A significant proportion, specifically more than 75%, of fatalities related to cardiovascular diseases (CVDs) occur in nations with lower economic status. Within these countries, elevated blood pressure is recognized as one of the primary risk factors contributing to the development of CVDs[4]. In the year 2016, India documented that 63% of the overall mortality cases were caused by non-communicable diseases (NCDs), with cardiovascular diseases (CVDs) accounting for 27% of these cases.

Cardiovascular diseases (CVDs) are responsible for 45% of mortality among those aged 40 to 69 years. Individuals who are susceptible to cardiovascular disease (CVD) may have elevated levels of blood pressure, glucose, and lipids, in addition to being overweight or obese [5]. According to the predictions of the World Health Organization, there would be an additional 24.5 million deaths in the year 2030.

In emerging countries, the lack of trained medical physicians and contemporary exam techniques has made detecting coronary heart diseases tough, slow, and complex. It is causing global issues . After reviewing the patient's medical history and symptoms, doctors recommend tests such as blood tests, ECG, coronary angiogram, exercise stress test, ultrasound, and nuclear cardiac stress test. In an ECG, the heart's electric impulses are read. Sticky dots on the palms, legs, and chest are attached to an ECG machine to record heartbeats as electrical impulses and print them on paper. Doctors prescribe this to diagnose coronary heart failure in irregularly beating hearts. MRI combines magnets and radio waves to take moving images [6, 16] of the heart on the computer and is recommended doctor An angiogram is done after a heart attack. A catheter is inserted into an artery in your wrist, arms, or legs and moved inside. A doctor uses a cardiac X-ray to see blocked arteries[7].

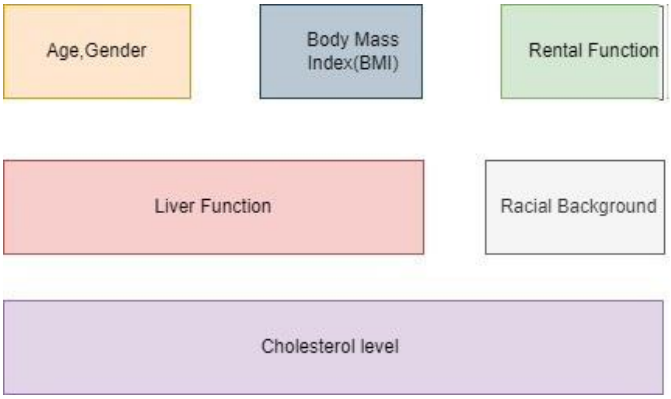

Fig 1: Factors influencing incidence in cardiovascular disease

Many early models for predicting heart disease utilizing AI computations were developed. Back vector machine, K-nearest neighbor, and Artificial Neural Network methods were developed early on to forecast whether or not a cardiac condition will occur. According to a survey, ANN-based models are usually utilized to diagnose cardiac disease, and previous work was not as accurate as the model that was created [7 ,10]. The necessity for diagnosing heart disease has prompted the development of several invasive clinical methods, such as a procedure called However, regardless its effectiveness, angiogram is associated with high costs and potential adverse consequences for individuals undergoing the procedure. The utilization of data mining techniques to identify cardiovascular disease (CVD) has been a subject of interest for several researchers, driven by the need to ensure safe and accurate diagnoses[11].

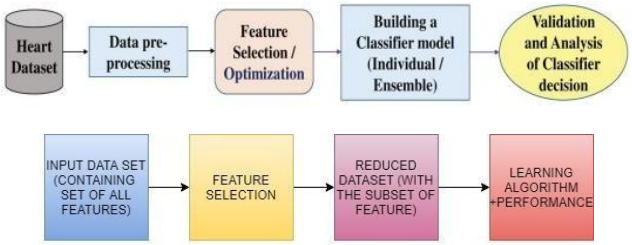

Fig 2: General block diagram of CVD detection model

2. LITERATURE REVIEW

There is a lot of related work in the domains that are directly related to this study. In the medical field, ANN has been introduced to produce the best accuracy prediction. This research presents a unique approach that utilizes machine learning approaches [20] to identify key variables, hence enhancing the accuracy of cardiovascular disease prediction. This research presents a prediction model that incorporates various combinations of variables and employs multiple established classification approaches. In this study, the researchers were able to get an improved level of performance in predicting heart disease using the hybrid random forest with a linear model (HRFLM). The accuracy level obtained 88.7% [1].

This study presents a comprehensive analysis of deep learning-based techniques that effectively integrate classification and feature augmentation tasks for the purpose of predicting cardiac issues in a dataset including patient records from five distinct medical institutes. The dataset has a total of 918 samples, each with a limited set of 11 clinical features. A novel architectural framework has been introduced, which integrates the Sparse Autoencoder with the Convolutional Classifier. Using this methodology, the researchers have attained a classification accuracy of 90.08%, signifying a notable enhancement of 4.4% when compared to the performance of traditional classifiers such as MLP or RF, which were trained on the identical dataset and subjected to identical circumstances [4].

This study presents a novel k-modes clustering technique utilizing Huang starting points, which has the potential to enhance the accuracy of categorization. The study utilizes the Random Forest (RF), Decision Tree Classifier (DT), Multilayer Perceptron (MP), and XGBoost (XGB) models. The GridSearchCV method is employed in order to optimize the results by fine-tuning the parameters of the model being utilized. The proposed methodology is evaluated using a dataset consisting of 70,000 instances from the Kaggle platform. The models underwent training using a data partitioned into an 80:20 ratio, resulting in the following accuracy measurements. The random forest model achieved an accuracy of 87.05% with cross-validation and 86.92% without cross-validation. The decision tree model achieved an accuracy of 86.37% with cross-validation and 86.53% without cross-validation. The XGBoost model achieved an accuracy of 86.87% with cross-validation and 87.02% without cross-validation. Lastly, the multilayer perceptron model achieved an accuracy of 87.28% with cross- validation and 86.94% without cross-validation. The AUC (area under the curve) values for the proposed models are as follows: The decision tree achieved an accuracy of 0.94, while XGBoost, random forest, and multilayer perceptron achieved accuracies of 0.95. The study's findings indicate that the multilayer perceptron, when combined with cross-validation, outperformed all other algorithms in terms of accuracy. The accuracy rate achieved was 87.28%, which was the highest among the tested models [2].

Based on the study findings, it has been shown that a classifier based on a stacking fusion model exhibits superior performance compared to separate models across all evaluation criteria. The aforementioned discovery implies that the practice of stacking models can effectively leverage the advantages offered by different types of models, resulting in enhanced predictive capabilities. The suggested stacking methodology provides enhanced predictive efficacy, heightened robustness, and augmented practicality for those with a heightened susceptibility to cardiovascular disease. Healthcare facilities have the potential to employ this data to discern individuals who possess a heightened susceptibility to cardiovascular illness and afterwards provide timely therapeutic interventions to mitigate this risk[3].

Subramani S. et al. suggested approach is assessed using the publicly available UCI heart-disease dataset, which consists of 1050 patients and 14 variables. The feature vector obtained from the heart- disease dataset was collected and utilized as input for a deep convolutional neural network (DCNN) in order to classify instances as either belonging to a healthy or cardiac disease class. In order to evaluate the effectiveness of the suggested approach, many performance measures were utilized, including accuracy, precision, recall, and the F1 measure. The validation accuracy of model achived to be 91.7%. The empirical findings demonstrate the efficacy of the suggested methodology within a real- world context [8, 9].

Arooj S. et al. proposed algorithm has the potential to assist medical practitioners in the early identification of potential heart disease cases, prior to the manifestation of symptoms. The primary objective of this study was to utilize a cutting-edge UCI library for the timely identification of cardiovascular

conditions. The dataset was first obtained from the UCI repository, consisting of a total of 76 cases. For the purpose of prediction, a subset of 14 examples was selected. Prior to conducting the training process, the data underwent preprocessing procedures. The preprocessed data was utilized in conjunction with the CNN algorithm for the purpose of predicting cardiac illnesses on the Google Collab platform. The performance metrics of accuracy, precision, recall, and F1 score were used to assess the suggested system. The system attained an accuracy of 91.71%, precision of 88.88%, recall of 82.75%, and F1 score of 85.70%[1].

This study conducted a comparative evaluation of four machine learning algorithms for the prediction of cardiac disease, obtaining favorable outcomes. The present analysis has demonstrated that machine learning methodologies have exhibited superior performance. The utilization of data pre-processing resulted in superior performance of XGBoost in the machine learning approach applied to the dataset of 13 characteristics. The XGBoost model yielded the greatest training and test scores, reaching 91% and 89% respectively. XGBoost achieved comparable outcomes, attaining an accuracy rate of 92% and an AUC score of 0.94 [6]..

3. DATA SET

Heart Disease Datasets: This section provides an overview of datasets commonly used in the reviewed publications. The most popular dataset used by the researchers is the Cleveland heart disease dataset obtained from the online repository of the University of California, Irvine (UCI) for machine learning[13]. It is comprised of 303 samples with 6 samples having missing values. The data, in its original form, have 76 features but all the published work is likely to refer to 13 features out of them and the other feature outlines the effect of the disease. In order to achieve correct results from machine learning algorithms, data is considered the first and most fundamental step. UCI machine learning repository is the source of the dataset that is used in the application[14].

Accurate results from machine learning algorithms depend on a wealth of data. UCI's machine learning repository provided the dataset used in this study. These five data sets cover the areas of Cleveland, Hungary, Switzerland, VA Long Beach, and Statlog. Combining them in this study allowed us to get more precise results. They have compiled a textbook with over 1190 examples and 14 distinguishing features from their database. There is a diagnostic input for each of the other 12 attributes, and a diagnostic output for each of the other 12 attributes. A patient's age, sex, BP at rest, BG at fasting, CP type, and EC tracings at rest were all present or nearly present in all medical records. Figure 3 displays the ranges of values assigned to a number of features, such as age, gender, cp, trestbps, chol, and more.

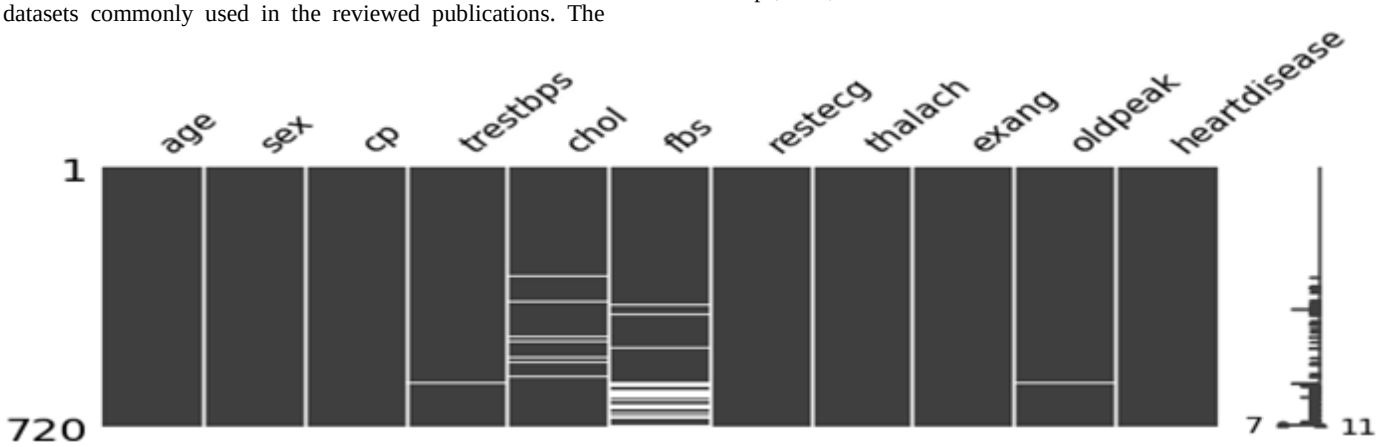

Fig 3: Attributes of Data set

4. PROPOSED MODEL

4.1 Data Preparation and Cleaning Techniques:

4.1.1 Choosing Features

To determine the most useful characteristics for data classification, machine learning can use a wide range of approaches. This is a bonus that contributes to the efficiency of the process. Here two main strategies: the Least Absolute Shrinkage and Selection Operator, and selecting relief features.

4.2 MACHINE LEARNING METHODS

More accurate classification results can be obtained using an ensemble technique because it combines multiple Decision Tree classifiers [15]. To boost the model's accuracy and precision, an ensemble of weak learners can be combined into a single strong learner. Figure 5 displays the three most common reasons for the discrepancy between the actual and predicted outcomes. All machine learning ensembles function in this way. The application of an ensemble method offers a potential solution to several of these problems.

4.3 PROPOSED METHOD

The objective of this study is to examine the efficacy of several machine learning algorithms in the prediction of cardiac disease. In order to accomplish this objective, here utilized a range of methodologies, such as K-nearest neighbors (KNN) [13], decision tree classifier, artificial neural network (ANN), and support vector machine (SVM) [14], in order to construct predictive models[17].In this article examine the machine learning approaches currently being used to create a reliable method for predicting heart illness[18].

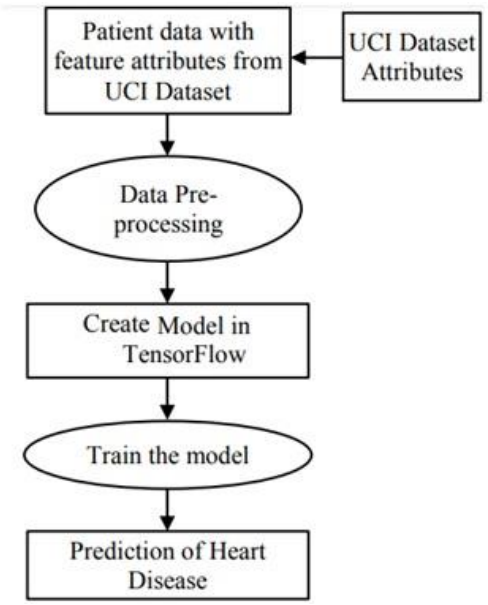

Figure 4: Implementation Techniques

Confusion Matrix: A classification model's performance is evaluated using a N x N confusion matrix, where N is the number of target classes. The matrix compares target values to machine learning model predictions[16].

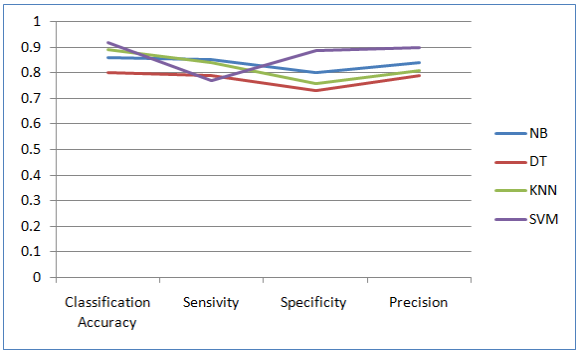

Fig 7: Proposed model accuracy

Table 2. Confusion Matrix Score

Method	Classification Accuracy	Sensitivity	Specificity	Precision
NB	0.86	0.85	0.8	0.84
DT	0.8	0.79	0.76	0.79
KNN	0.89	0.86	0.76	0.81
SVM	0.92	0.87	0.89	0.9

5. CONCLUSION

In this research, machine learning approaches[19] for identifying heart disease. Cardiovascular disease is a potentially fatal condition that impacts a significant global population annually. Therefore, the timely identification of cardiac disease holds potential advantages for both patients and healthcare practitioners, as it equips them with crucial data to mitigate mortality rates and reduce financial burdens. With the continuous growth of medical big data and the lowering costs of data storage, the utilization of machine learning algorithms has emerged as a significant factor in the processing and prediction of diseases based on medical data. By utilizing the Random Forest Classifier technique, a machine-learning model was successfully constructed. The model underwent training and testing using a dataset sourced from the UCI library. The dataset comprised 303 participants, all of whom were assigned labels indicating their respective medical conditions. This encompassed individuals diagnosedwith heart illness as well as those without any known cardiac ailments. Upon completion of training and subsequent testing, an accuracy of 92.4% was attained using the default hyperparameter configuration. During experimentation with classifiers, achieved the default value provided the maximum level of accuracy.

6. REFERENCES

[1] Arooj S, Rehman SU, Imran A, Almuhaimeed A, Alzahrani AK, Alzahrani A. A Deep Convolutional Neural Network for the Early Detection of Heart Disease. Biomedicines. 2022 Nov 3;10(11):2796. doi: 10.3390/biomedicines10112796. PMID: 36359317; PMCID: PMC9687844.

[2] Bhatt, C.M.; Patel, P.; Ghetia, T.; Mazzeo, P.L. Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms 2023, 16, 88. https://doi.org/10.3390/a16020088

[3] D. Tay, C. L. Poh, E. Van Reeth and R. I. Kitney, "The Effect of Sample Age and Prediction Resolution on Myocardial Infarction Risk Prediction," in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 3, pp. 1178-1185, May 2015. doi: 10.1109/JBHI.2014.2330898

[4] Dwivedi, A.K. Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Comput & Applic 29, 685–693 (2018). https://doi.org/10.1007/s00521-016-2604- 1

[5] García-Ordás, M.T., Bayón-Gutiérrez, M., Benavides, C. et al. Heart disease risk prediction using deep learning techniques with feature augmentation. Multimed Tools Appl 82, 31759–31773 (2023). https://doi.org/10.1007/s11042-023-14817-z

[6] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in IEEE Access, vol. 8, pp. 107562- 107582, 2020. doi: 10.1109/ACCESS.2020.3001149

[7] Nandal N, Goel L and TANWAR R. Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis [version 1; peer review: 1 approved]. F1000Research 2022, 11:1126 -https://doi.org/10.12688/f1000research.123776.1

[8] Ramprakash, P.; Sarumathi, R.; Mowriya, R.; Nithyavishnupriya, S. (2020). [IEEE 2020 International Conference on Inventive Computation Technologies (ICICT) - Coimbatore, India (2020.2.26-2020.2.28)] 2020 International Conference on Inventive Computation Technologies (ICICT) - Heart Disease Prediction Using Deep Neural Network. , (), 666–670. doi:10.1109/ICICT48043.2020.9112443

[9] Subramani S., Varshney N.et al.(2023) Cardiovascular diseases prediction by machine learning incorporation with deep learning, Frontiers in Medicine VOLUME 10, 10.3389/fmed.2023.1150933, 2296-858X

[10] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019. doi: 10.1109/ACCESS.2019.2923707

[11] Y. Meng, W. Speier, C. Shufelt, S. Joung, J. E. V. Eyk et al. (2019). , "A machine learning approach to classifying self-reported health status in a cohort of patients with heart disease using activity tracker data," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 3, pp. 878–884.

[12] Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian M.W. Quinn, Mohammad Ali Moni (2021) Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, Computers in Biology and Medicine, Volume 136, 104672, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2021.104672.

[13] C. Beulah Christalin Latha, S. Carolin Jeeva (2019) Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, Informatics in Medicine Unlocked, Volume 16, 100203, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2019.100203.

[14] Shadi Tajmiri, Ebrahim Azimi, Mohammad Raouf Hosseini, Yousef Azimi, (2020) Evolving multilayer perceptron, and factorial design for modelling and optimization of dye decomposition by bio-synthetized nano CdS-diatomite composite, Environmental Research, Volume 182, 108997, ISSN 0013-9351, https://doi.org/10.1016/j.envres.2019.108997.

[15] Yousef Azimi, Mohammad Talaeian, Hamid Sarkheil, Rana Hashemi, Ravanbakhsh Shirdam (2022) Developing an evolving multi-layer perceptron network by genetic algorithm to predict full- scale municipal wastewater treatment plant effluent, Journal of Environmental Chemical Engineering, Volume 10, Issue 5, 108398, ISSN 2213-3437, https://doi.org/10.1016/j.jece.2022.108398.

[16] D. Jeni Jeba Seeli & K. K. Thanammal. (2023) Quantitative Analysis of Gradient Descent Algorithm using scaling methods for improving the prediction process based on Artificial Neural Network. Multimedia Tools and Applications.

[17] S. R. Lavanya & R. Mallika. (2022) AMCGWO : An enhanced feature selection based on swarm optimization for effective disease prediction. Journal of Discrete Mathematical Sciences and Cryptography 25:3, pages 635-647.

[18] Mishra Jayant & Goyal Sachin. (2022). An effective automatic traffic sign classification and recognition deep convolutional networks. Multimedia Tools and Applications. 81. 10.1007/s11042- 022-12531-w.

[19] Yanji Qu, Xinlei Deng, Shao Lin, Fengzhen Han, Howard H. Chang, Yanqiu Ou, Zhiqiang Nie, Jinzhuang Mai, Ximeng Wang, Xiangmin Gao, Yong Wu, Jimei Chen, Jian Zhuang, Ian Ryan & Xiaoqing Liu. (2022) Using Innovative Machine Learning Methods to Screen and Identify Predictors of Congenital Heart Diseases. Frontiers in Cardiovascular Medicine 8.

[20] Zakir Hussain & Malaya Dutta Borah. (2020) Birth weight prediction of new born baby with application of machine learning techniques on features of mother. Journal of Statistics and Management Systems 23:6, pages 1079-1091.