International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 10 ǁ October 2024 ǁ PP. 01-08

Predicting Psychological Disorder in Parents of Children With Nephrotic Syndrome: A Comparative Study Of Six (6) Machine Learning Classifiers

Adelani Ayobami[1*], Omolayo Clement[2*], Adebayo O.T [3*], Okunlola Yusuf[4*]
[1,2]*Federal University of Technology, Akure, Ondo State, Nigeria* [3]*St Lawrence College, Kingston ON, Canada*
[4]*University of Ibadan, Oyo State, Nigeria.*
Corresponding Author: Adelani Ayobami ayobamiadelani101@gmail.com

ABSTRACT
The diagnosis of nephrotic syndrome in a child can significantly impact the psychological well-being of parents. Identifying parents at risk of developing psychological disorders is crucial for providing timely support and interventions.
This research project aims to predict the likelihood of psychological disorders in parents of children with Nephrotic Syndrome using machine learning techniques. A comparative study of six classifiers is conducted on a dataset consisting of relevant features and psychological assessments. The performances of the decision tree classifier, random forest classifier, support vector classifier, cat boost classifier, K-nearest neighbors classifier, and logistic regression are evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC. The findings have implications for early identification and targeted support, enhancing the well-being of parents and children affected by nephrotic syndrome. This study contributes to predictive healthcare analytics and highlights the potential of machine learning to address the psychological impact of the condition. The results from the study show CatBoost Classifier and Decision Tree Classifier have AUC values of 1.0, followed by Random Forest Classifier and Logistic Regression with AUC values of 0.94. The two least performed algorithms are K-Nearest Neighbors and Support Vector Classifier, with AUC values of 0.72 and 0.56, respectively.
Keywords*: Nephrotic Syndrome (NS), Machine Learning Classifiers (MLC), Psychological Disorders, Mental Health, Paediatrics*

Date of Submission: 01-10-2024 Date of acceptance: 11-10-2024

I. INTRODUCTION

The diagnosis of Nephrotic Syndrome in a child can be a distressing experience for parents, as they navigate the challenges of managing their child's health and coping with the emotional and psychological impact of the condition. The psychological well-being of parents plays a crucial role in providing effective support and care for their child. Identifying parents at risk of developing psychological disorders is essential for early intervention and targeted support [2].

The objective of this research project is to explore the predictive capabilities of machine learning techniques in determining the likelihood of psychological disorders in parents of children diagnosed with Nephrotic Syndrome. By leveraging the power of machine learning classifiers, we aim to develop an accurate and reliable prediction model that can identify parents who may be at risk. The comparative study will involve evaluating and comparing the performance of six machine learning classifiers, namely Logistic Regression, Support Vector Classifier, Random Forest, CatBoost Classifier, K-Nearest neighbors and Decision Tree. These classifiers will be trained and tested using a dataset consisting of relevant features and psychological assessments of parents.

The research methodology will encompass data collection from a previous study on parents of children diagnosed with Nephrotic Syndrome, ensuring ethical considerations and confidentiality. The dataset includes demographic information, medical history, and standardized psychological assessments to capture various dimensions of psychological well-being. The extracted data will undergo pre-processing, including removal of outliers and filling of missing values, to ensure the quality and reliability of the input. The machine learning classifiers will be trained on the pre-processed dataset using appropriate algorithms, and their performance will be evaluated using various metrics such as accuracy, precision, recall, F1-score, and AUC. The comparison of the classifiers will provide insights into their effectiveness in predicting the likelihood of psychological disorders in parents.

The findings of this research project will have significant implications for healthcare professionals and support systems involved in the care of families affected by Nephrotic Syndrome. The development of an

accurate prediction model can enable early identification of parents at risk, facilitating timely intervention, counselling, and mental health support. By providing targeted assistance to parents experiencing psychological challenges, we can enhance the overall well-being of both parents and children affected by Nephrotic Syndrome.

Overall, this research aims to leverage machine learning techniques to address the critical issue of predicting psychological disorders in parents of children diagnosed with Nephrotic Syndrome. The potential impact of this study lies in its ability to contribute to personalized care, improved mental health support, and enhanced overall outcomes for families affected by this chronic condition.

Aim:

1. To develop a model that can predict psychological disorder in parents of children with Nephrotic Syndrome.
2. To compare the predictions of psychological disorder in parents of children with NS using six (6) Machine Learning classifiers.

II. RELATED WORKS

[3] examined the impact of Pediatric Nephrotic Syndrome (NS) on parents' health-related quality of life (HRQOL) and family functioning using the PedsQL 4.0 Family Impact Module (FIM). A cross-sectional study comparing parents of children with chronic NS (n=61) to a healthy control group (n=72) revealed significantly lower scores in all categories of the PedsQL TM FIM questionnaire among parents of NS children. Female gender of the affected child was identified as a risk factor for poor family functioning, and the presence of serious complications predicted lower HRQOL and family impact scores. These findings highlight the substantial disruption of HRQOL and family functioning caused by NS in parents, necessitating further research and interventions to address these challenges.

In the paper [4], the research aimed to assess the level of parental burden and psychological distress among parents of children with nephrotic syndrome. The study was conducted in a tertiary care hospital in Bhubaneswar, Odisha, India, using a cross-sectional design. The sample consisted of 62 parents of children with nephrotic syndrome, and data were collected through structured questionnaires measuring parental burden and psychological distress. The study found a strong and positive correlation between parental burden and psychological distress. The research contributes to understanding the challenges faced by parents in managing a child's nephrotic syndrome and the associated psychological impact.

[5] conducted a comparative survey on machine learning approaches for the prediction of mental health issues in adolescents. The study aimed to explore the efficacy of various machine learning algorithms and statistical analysis tools in early detecting and predicting mental health disorders among adolescents. The survey reviewed 22 previous research papers categorized into different mental health disorders, including depression and anxiety, suicidal prevalence, autism spectrum disorder (ASD), and substance abuse among adolescents. The authors compared the performance of machine learning prediction models based on accuracy, with CNN models, Random Forest, and XGBoost generally showing better results than other models. The study emphasized that machine learning algorithms can be effective in classifying and predicting high-risk factors for mental health issues among adolescents. The research primarily focused on centralized research in Pakistan, where SPSS and other tools were commonly used for data analysis. The findings highlight the potential of machine learning techniques in improving early detection and intervention for mental health issues in adolescents.

[6] focused on the early prediction of chronic kidney disease (CKD) in adolescents using machine learning. The research was conducted at Dayananda Sagar University in Bangalore. The study aims to utilize machine learning algorithms to predict the occurrence of CKD based on various symptoms, enabling early diagnosis and timely treatment for improved health and quality of life. The paper explores the association of data parameters and target class attributes through predictive analytics, analyzing the prediction skills of different machine learning algorithms. The research highlights the potential of machine learning in early CKD prediction and emphasizes its significance in the context of adolescent healthcare.

In a recent study by [7], machine learning models were developed to predict steroid-resistant nephrotic syndrome (SRNS). The study utilized a combination of penalized regression and nonparametric screening methods to select 26 informative clinical variables. A support vector machine (SVM) model was built using these variables, achieving a high accuracy of 95.2% in leave-one-out cross-validation. A reduced SVM model with eight clinical variables demonstrated a validation accuracy of 94.0%, sensitivity of 90.0%, and specificity of 96.7%. Notably, the inclusion of vinculin autoantibody highlighted its linear relationship with steroid responsiveness. The model provides a valuable tool for evaluating and selecting appropriate treatment methods for nonhereditary SRNS.

[8] explored the application of machine learning techniques for predicting mental health problems in children. The study compares the performance of eight machine learning techniques on a dataset of sixty cases, focusing on accuracy in diagnosing five basic mental health problems. Results indicate that the Multilayer

Perceptron, Multiclass Classifier, and LAD Tree classifiers exhibit higher accuracy, with minimal variation between the full attribute set and selected attribute set. This research contributes to early detection and intervention efforts in children's mental health through the utilization of machine learning techniques.

In the study [9], machine learning algorithms were trained and evaluated to predict stunting among under-five children in Zambia using the Zambia Demographic Health Survey (ZDHS) dataset. Logistic regression, Random Forest, SV classification, XG Boost, and Naïve Bayes algorithms were applied, and the best-performing algorithm was found to be Random Forest with an accuracy score of 79% in testing data. The study highlighted the potential of machine learning models in aiding the timely diagnosis and prevention of stunting among children in Zambia.

Drawing upon the insights gained from the literature review of relevant studies, this comparative study aims to explore the efficacy of five machine learning classifiers in predicting psychological disorders in parents of children with nephrotic syndrome. By comparing the performance of different machine learning algorithms, including XGBoost Classifier, RandomForest Classifier, Support vector classifier, Cat boost classifier, and Logical regression, the study seeks to identify the most accurate and reliable classifier for predicting psychological disorder in this specific population. The research intends to contribute to the understanding of the predictive capabilities of machine learning in assessing the psychological well-being of parents dealing with nephrotic syndrome, enabling early detection and intervention for improved support and intervention strategies

III. METHODOLOGY

DATA EXPLORATION

Since this project aims to predict the psychological disorders of caregivers of children with NS, the dataset was extracted from an observational study conducted in the paediatric nephrology units of two public-funded hospitals in Lagos State, southwest Nigeria, over 7 years (February 2012 to April 2019). The dataset consists of various features that include: child characteristics (the demographic data, steroid status, duration of hospitalization, and number of relapses); and caregiver's characteristics (the relationship to the child, the demographic data, and their social and economic status).

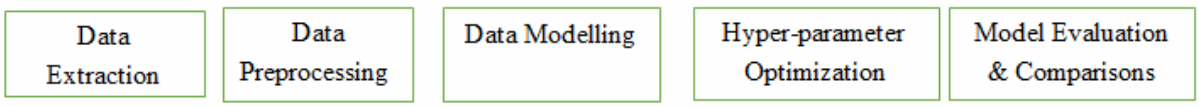
Table 1. Features Description

Variables	Description	Value Type
mZarit	It is the target variable with binary data (0,1)	1 – means no psychological distress 2 – means presence of psychological distress
Gender	It is a demographic variable with two values, male and female.	m, f
Ageyr	It is a demographic variable, the age of the child in years	Continuous data
Caregiverage	It is a demographic variable, the age of the caregiver in years	Continuous data
Cargivergender	It is a demographic variable with two values, male and female for caregivers	m,f
MaritalStat	It is a categorical variable that indicates caregiver's marital status.	Married, single, divorced, separated, widowed
CaregiverRel	It is a categorical variable that indicates caregivers' religion.	Islam, Christianity
rTribe	It is a categorical variable that indicates caregivers' ethnicity	Yoruba, Ibo, Others
SES	It is a categorical variable that indicates caregivers' social economic status.	Low, middle, high
rHosp6month	It is a categorical variable that indicates if the child was hospitalized in the last 6month.	No, yes
rSour	It is a categorical variable that indicates if caregivers relied on non-family source to pay for healthcare	Fam, Others
SteroidStat	It is a categorical variable that indicates the child's steroid status.	Sensitive, resistant
Boys	It is an ordinal value that tells the numbers of boys in the family.	Ordinal data
Girls	It is an ordinal value that tells the numbers of	Ordinal data

	girls in the family.	
Children	It is an ordinal value that tells the numbers of children in the family.	Ordinal data
GHQ_Score	It is a tool that describes the psychological stress or minor mental health morbidity focusing mainly on symptoms of depression, anxiety. The possible range score of GHQ is between 0 and 12.	Ordinal data
Zar_6	It is a reliable instrument that test caregivers' perception of their psychological, social, physical health and their financial status. The values of Zar_6 >= 6 indicates significant caregiver burden.	Ordinal data
Duration	It is a continuous variable that describe the duration in months the child was hospitalized.	Continuous data
Groups	It is a categorical variable that indicates if the child was Steroid Sensitive, Frequently Relapse, and Steroid Resistant.	SSNS, FRSD, SRNS
ImpairedHealth	It is a categorical variable that indicates if the child had any impairment.	No, yes
rRelapse	It is a categorical variable that indicates if the child was currently on relapse.	No, yes

The dataset used in this research work was extracted from a published paper by [1]. The data was collected from two tertiary teaching hospitals in Lagos, Nigeria.

The framework of this project is shown in the figure below:



Data Pre-Processing

This crucial stage involved a number of operations, including the elimination of superfluous columns, handling of missing values, the elimination of duplicate rows, and the encoding of categorical variables into numerical values prior to the machine learning algorithms being applied to the cleaned dataset, which was split into 70% training and 30% validation. The dataset used in this study contain 172 rows and 21 columns.

Data Modelling

In this project, six machine learning algorithms were utilized that include Logistic Regression, Support Vector Classifier, CatBoost Classifier, Random Forest, Decision Tree, and KNN.

1. Logistic Regression
2. Support Vector Classifier
3. Random Forest
4. CatBoost Classifier
5. Decision Tree
6. KNN

Machine Learning Classifiers

Supervised machine learning employs classification algorithms as predictive techniques. This entails utilizing a set of labeled instances distributed across at least two classes (attributes) of objects and forecasting the value of a specific categorical class (attribute) based on the values of other predictive attributes. Through the analysis of attribute values, the classification algorithm uncovers relationships, ensuring precise prediction outcomes. Machine learning classifiers have been successfully used in medical research. Classifiers widely used include Logistic Regression, Support Vector Classifier (SVC), Random Forest, CatBoost Classifier, Decision Tree, and K-Nearest Neighbors (KNN). The following section briefly presents the classification algorithms used in the study.

Model Evaluation

This is yet another important section of the paper. This section quantified the performance of each categorization model by taking into account three essential evaluation variables. The confusion matrix served as the basis for the measurements, which included the F1-score, accuracy, and Receiver Operating Characteristics (ROC) curve.

Table 2
Confusion Matrix

	Predicted Values	
Actual Values	Negative (0)	Positive (1)
Negative (0)	True Negative (TN)	False Positive (FP)
Positive (1)	False Negative (FN)	True Positive (TP)

IV. RESULTS

This section of the paper shows the results of each machine learning algorithm by displaying their accuracy, F1 score, and AUC value from the receiver operating characteristics curve.

Table 3
Results of performance for models without hyperparameter optimization

Classifier	Confusion Matrix	Train Accuracy (%)	Test Accuracy (%)	F1-Score	AUC Value
Logistic Regression	[43 0] [1 8]	100	98.08	0.94	0.94
Support Vector Classifier	[43 0] [8 1]	82.50	84.62	0.20	0.56
Random Forest	[43 0] [1 8]	100	98.08	0.94	0.94
CatBoost Classifier	[43 0] [0 9]	100	100	1.00	1.00
Decision Tree	[43 0] [0 9]	100	100	1.00	1.00
KNN	[43 0] [5 4]	93.33	90.38	0.62	0.72

As shown in the table above, two models had the same accuracy of 100%, which includes the CatBoost classifier and the Decision Tree classifier; also, two models had the same accuracy of 98.08%, which includes the Logistic Regression classifier and the Random Forest classifier; and the lowest amongst all, with an accuracy of 84.62%, is the Support Vector Classifier. The same applies to the F1-Score and their AUC values. Therefore, the best models with high accuracy, F1-Score, and AUC that performed excellently without flaws are the CatBoost Classifier and Decision Tree.

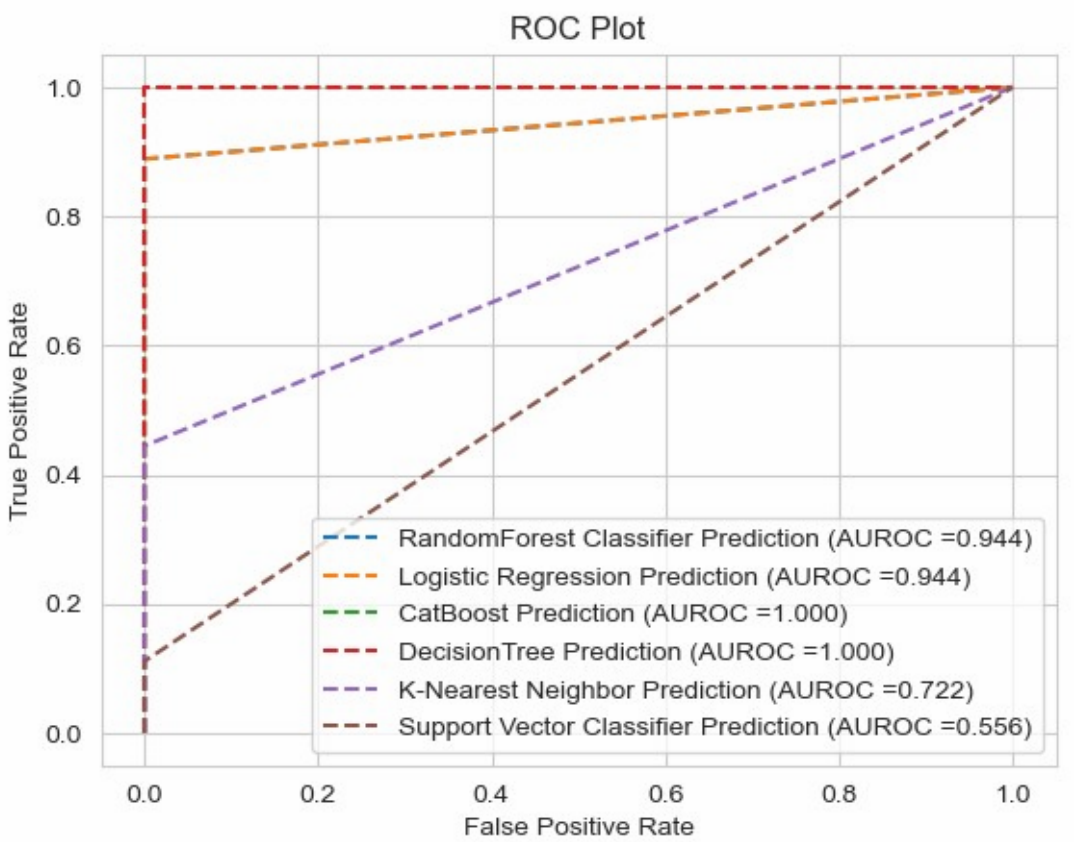


Figure 1. ROC Curve for Baseline models

The curve is the ROC plot for the six models without hyperparameter optimization. The best models that performed flawlessly with an AUROC score of 1.000 were the CatBoost classifier and the Decision Tree classifier, while the least model with an AUROC value of 0.556 was the Support Vector classifier.

Table 4. Results of performance for models with hyperparameter optimization.

Classifier	Confusion Matrix	Train Accuracy (%)	Test Accuracy (%)	F1-Score	AUC Value
Tuned-Logistic Regression	[43 0] [2 7]	100	96.15	0.88	0.89
Tuned-Support Vector Classifier	[43 0] [5 4]	84.17	90.38	0.62	0.72
Tuned-Rando m Forest	[43 0] [1 8]	100	98.08	0.94	0.89
Tuned-CatBoost Classifier	[43 0] [0 9]	100	100	1.00	1.00
Tuned-Decision Tree	[43 0] [0 9]	100	100	1.00	1.00
Tuned-KNN	[43 0] [4 5]	96.67	92.31	0.71	0.78

As shown in Table 4 above, the best models with perfect accuracies of 100% were still Tuned-CatBoost Classifiers and Tuned-Decision Tree Classifiers. The duo models perfectly predicted the y-label and mZarit scores (0, 1) after their parameters were tuned. The least tuned model on the dataset was the Tuned-Support Vector Classifier.

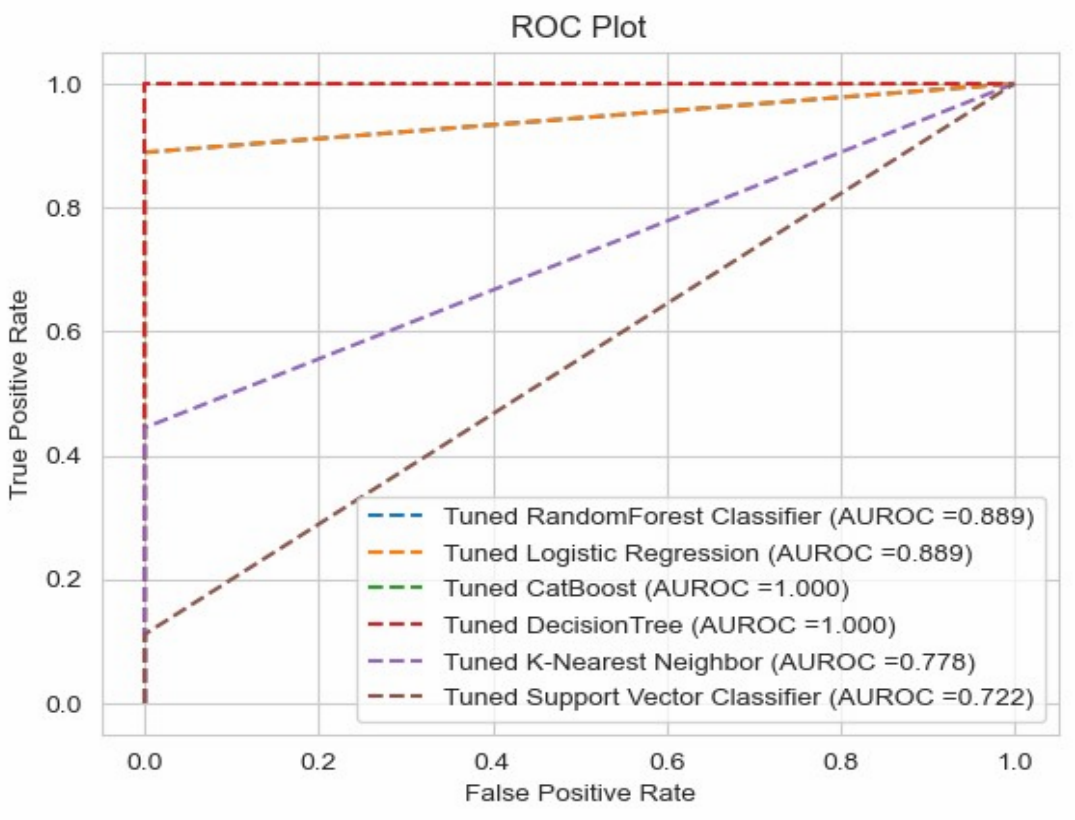


Figure 2. ROC Curve for Tuned models

The curve above is the ROC curve for the six models with hyperparameter optimization. The best models with AUROC (Area Under Receiver Operating Characteristics) of 1.000 were still Tuned-CatBoost Classifier and Tuned Decision-Tree. The least model with an AUROC of 0.722 was the Tuned Support Vector Classifier.

V. DISCUSSION

The implementation of machine learning to create a solution that would help physicians address the mental stability of caregivers due to the diagnosis of their child's health led to the discovery of this study. This ideology motivated the authors of this research to build predictive models that would help to predict mental burdens in parents of children with nephrotic systems and also guide physicians on how to identify parents suffering from psychological disorders to prevent unprecedented situations. This study will be instrumental in medical diagnosis to prevent lives from health conditions that would shorten their lifespan. The study implemented six supervised machine learning classifiers to make accurate predictions and also compared the performances of these algorithms. The framework for this study follows this pattern, which includes data extraction from a published work, data cleaning, data modeling, hyperparameter optimization, model evaluation, and comparisons. The data is small, with 172 rows and 21 features that show the attributes of each caregiver. The cleaned data was fitted into the predictive models to make reliable predictions that will aid physicians in making adequate recommendations for the proper healthcare and wellbeing of caregivers of children with nephrotic syndrome. The result of this study shows that the decision tree and catboost performed excellently amongst the six classifiers, while the support vector classifier is the least popular amongst all.

VI. CONCLUSION

The aim of the study was accomplished with the implementation of machine learning to build robust models that will aid in making medical diagnoses for caregivers of children suffering from nephrotic syndrome. The burden of taking care of these children might have a strong effect on the mental health of their caregivers, which could result in untimely death or suicide if not properly addressed early by physicians. Six classifiers that include Decision Tree, Support Vector Machine, Logistic Regression, CatBoost, RandomForest, and K-Nearest Neighbors were trained to make accurate predictions, with four algorithms (CatBoost, Decision Tree, RandomForest, and Logistic Regression) predicting accurately while the other two algorithms (K-Nearest Neighbors and Support Vector Machine) still have more false negatives in their predictions.

One of the key limitations of the study is the small amount of dataset used for training the models, where it's a game of chance to generalize the accuracy of the trained models on a large dataset. The authors would recommend future works consider using a large dataset.

Since supervised machine learning algorithms were used in this study, future works could also implement deep learning approaches for comparing the performances of the predictive models to ascertain the reliability of this research. This research work can also be extended by software developers to build a web platform that would enable simple interaction with the models to enable user-centric accessibility for physicians.

REFERENCES

[1]. Esezobor, C.I, Solarin, A.U., Olagunju, A.T. (2020). Significant Burden and Psychological Distress Among Caregivers of Children with Nephrotic Syndrome: A Cross-Sectional Study, Canadian Journal of Kidney Health and Disease, Vol.7 1-8, DOI: 10.117/2054358119898016.

[2]. NHS Editorial Team (2022). Nephrotic Syndrome in Children. https://www.nhs.uk/conditions/nephrotic-syndrome/. Retrieved 12th July, 2023.

[3]. Mishra, K., Ramachandran, S., Firdaus, S., & Rath, B. (2015). The impact of pediatric nephrotic syndrome on parents' health-related quality of life and family functioning: An assessment made by the PedsQL 4.0 Family Impact Module. Saudi Journal of Kidney Diseases and Transplantation, 26(2), 285-292.

[4]. Sen, P., Das, N., Sahoo, P., & Mohanty, J. (2022). Parental burden and psychological distress among parents of children with nephrotic syndrome: A cross-sectional study. Journal of pharmaceutical negative results, Vol. 13 special issue 09 (2022), Page range. DOI:10.47750/pnr.2022.13.S09.824.

[5]. Haroon, K., Minhas, S., Sabahat, N., & Nassrani, S. (2023). Machine Learning Approaches for Prediction of Mental Health Issues in Adolescents: A Comparative Survey. VFAST Transactions on Software Engineering, 11(1), 37. Retrieved from http://vfast.org/journals/index.php/VTSE

[6]. Stella, A., & Kumari, V. P. (2020). Early Prediction of Chronic Kidney Disease in Adolescents using Machine Learning. International Journal of Advanced Research in Computer and Communication Engineering, 9(7), 31. DOI: 10.17148/IJARCCE.2020.9706

[7]. Ye, Q., Li, Y., Liu, H., Mao, J., & Jiang, H. (2023). Machine learning models for predicting steroid-resistant nephrotic syndrome. Frontiers in Immunology, 14, Article 1090241. https://doi.org/10.3389/fimmu.2023.1090241

[8]. M.R., S., & Poorna, B. (2016). Prediction of mental health problems among children using machine learning techniques. International Journal of Advanced Computer Science and Applications, 7(1). Retrieved from https://www.researchgate.net/publication/292944128_Prediction_of_Mental_Health_Problems_Among_Children_Using_Machine_Learning_Techniques

[9]. Chilyabanyama, O.N., Chilengi, R., Simuyandi, M., Chisenga, C.C., Chirwa, M., Hamusonde, K., Saroj, R.K., Iqbal, N.T., Ngaruye, I., & Bosomprah, S. (2022). Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia. Children, 9(7), 1082. https://doi.org/10.3390/children9071082

[10]. Hossain, R., and Timmer, D. (2021). Machine Learning Model Optimization with Hyper Parameter Tuning Approach. Global Journal of Computer Science and Technology: D Neural & Artificial Intelligence. 21 (2)

[11]. Lim, S.T, Khaw, K.W., Chew, X.Y., Yeong, W.C. (2023). Prediction of Thyroid Disease using Machine Learning Approaches and Featurewiz Selection. Journal of Telecommunication, Electronic and Computer Engineering. 15 (3)