Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 4 (2025) 277

African Scientific Reports

Enhanced machine learning model for classification of the impact of technostress in the COVID and post-COVID era

Gabriel James[a,*], Anietie Ekong[b], Aloysius Akpanobong[c], Enefiok Etuk[d], Saviour Inyang[a], Samuel Oyong[a], Ifeoma Ohaeri[a], Chikodili Orazulume[e], Peace Okafor[f]

[a]*Department of Computing, Topfaith University, Mkpatak, Nigeria*
[b]*Department of Computer Science, Akwa Ibom State University, Ikot Akpaden, Nigeria*
[c]*Department of Computer Networks, Faculty of Science and Information Technology, UPM, Serdang, Malaysia*
[d]*Department of Computer Science, Michael Okpara University of Agriculture, Umudike, Nigeria*
[e]*Department of Electrical Electronics Engineering, Topfaith University, Mkpatak, Nigeria*
[f]*Department of State Service, Ebonyi State Command, Abakaliki, Nigeria*

Abstract

The global crisis caused by the coronavirus outbreak and other diseases has significantly changed daily life, work, and education, forcing individuals and organizations to adapt to evolving virtual environments. These challenges have led to conditions induced by an inability to process information effectively with computer technologies. This study models a system that employs a Random Forest algorithm for prediction and classification, using age, gender, hours spent, and technological experience as parameters to categorize stress into high, moderate, and low levels. Data were collected via a questionnaire during the COVID-19 and post-COVID eras using a non-probabilistic sample of knowledgeable respondents. The model achieved 90% accuracy, demonstrating its prediction efficiency. Additionally, an interactive user interface was developed to facilitate real-time evaluation of technostress's impact on technology use. This work contributes a novel machine learning framework for technostress assessment, providing a practical tool for organizations and policymakers to better understand and mitigate technology-induced stress.

DOI:10.46481/asr.2025.4.1.277

Keywords: COVID-19 era, Technostress, Machine learning models, Machine learning, AI, Deep learning

Article History :
Received: 12 January 2025
Received in revised form: 05 February 2025
Accepted for publication: 12 March 2025
Published: 03 April 2025

1. Introduction

The post-COVID-19 era has claimed a large number of lives around the world and poses an unprecedented threat to public health, food systems, and the workplace [1]. According to WHO reports, the pandemic's economic and social

*Corresponding author Tel. No.: +234-902-137-9947.
Email address: g.james@topfaith.edu.ng (Gabriel James)

impacts are catastrophic: Tens of millions of people are at risk of experiencing extreme poverty, and by the end of the year, the number of undernourished people—currently estimated to be over 690 million—may have increased to 132 million. The COVID-19 crisis has resulted in years of change in how businesses in all sectors and regions operate. Also, due to social conventions and state-wide lockdowns, the COVID-19 pandemic area has inevitably increased the usage of digital technologies. People and organizations around the world have had to adapt to new distancing work and lifestyles. The influence of technological advancement in this COVID-19 era has affected people's lives and states in a few industrialized and technologically advanced countries. Since the outbreak of the COVID-19 pandemic, many businesses and organizations have encouraged their employees to work from home to avoid direct contact and mingling with co-workers. Distance learning among students and teachers has improved thanks to technological advancements. Because of the rising number of people afflicted with the coronavirus, several governments have ordered that all in-person learning programs in institutes be halted to assist in stopping the spread of the virus. Following the global outbreak of COVID-19, technological advancements have turned online shopping from a once-in-a-while activity to a must-do worldwide to reduce human movement and thus control the spread of coronavirus. As a consequence of improvements in technology, several nations have embraced the usage of soft money or contactless payments to pay for any services.

Governments at the regional, local, and international levels are frantically trying to resolve the COVID-19 crisis. The government has had to ensure that all available technology is being deployed to combat the pandemic and deal with a variety of COVID-19-related difficulties during this time. As a result of the COVID-19 crisis, this essay will explore the application of technical methods and the growth of technology in several industries. It also discusses how the government and health groups have implemented new rules in an attempt to halt the spread of coronavirus. As a result of these new restrictions, lockdowns and social distancing measures, for instance, have led to technological innovation and new ways of communicating with the government, corporations, and individuals. Expanded web-based shopping, mechanical conveyance frameworks, the coming of computerized and contactless instalment frameworks, remote working, the significance of innovation in far-off learning, Telehealth, 3D printing, and online diversion are only a few of these changes. A few countries have embraced these technological advancements throughout this pandemic, albeit with certain restrictions in specific developing and impoverished nations helping people to adapt to COVID-19.

The COVID-19 outbreak has had a major impact on the telecommunication sector. It increased the number of people using media apps for virtual meetings, teleworking, virtual education, and social relationships. Too much time spent in front of a smartphone, desktop, or smart device can lead to stress and worry. Telecommunication-related mental health issues can exacerbate other pressures like isolation and lockdown, leading to exhaustion and burnout. This review looks at the impact of the COVID-19 epidemic on communication and education. The relationship between the long-term use of electronic devices and mental well-being is also examined [2]. Finally, coping strategies are proposed to help mitigate pandemic Tele-burdens. The term "hurtful psychological condition generated by the use of ICTs or the danger of ICTs" refers to techno-stress, which Tarafda defines as the result of an imbalance between the resources and needs related to using ICTs, which leads to severe discomfort, psychosomatic activation, and the formation of negative attitudes toward ICTs. Cuervo *et al.* describe techno-stress as the negative impact that technology has, either directly or indirectly, on attitudes, beliefs, actions, or body physiology. Conversely, Cuervo *et al.* describe techno-stress from a medical perspective [3]. Techno-stress is a reactive sickness induced by an inability to deal with new information healthily with computer technologies [2].

Techno-stress is still not well understood because it is a recent phenomenon. According to Educational Research, techno-stress is "a modern sickness of adoption caused by an incapacity to cope with new technology healthily." One of two ways that techno-stress manifests is that people may either find it difficult to embrace computer technology or become overly identified with and dependent on it. Even though modern society is surrounded by superior technology, this fact might be problematic. Many people in the hyper-connected culture and concerned about techno-stress, or the negative effects that using computers has on people. Techno-stress undoubtedly affects how people think, feel, and interact with their environment, even if a scientific study on the subject is still in its early stages. This research work aims to develop a hybrid machine-learning model for the classification of the impact of techno-stress in COVID-19 and post-ERA. It will be relevant due to the embodiment of real-life and scientific problems. It shall showcase effective methods for predicting and classification of techno-stress data; and enhance and create awareness about the effects of techno-stress; with the number of participants, the data acquired could be used for medical and psychological purposes for the examination of techno-stressed persons and also serve as a tool to further research for medical for academic purposes. The work is focused on the classification of the impact of technostress using a machine learning algorithm. The analysis will be carried out using a scientific questionnaire for data collection.

Computer-associated stress is a complex psychological state that occurs when using a computer [3]. It can be a temporary or ongoing state of worry that is directly tied to computer use. Excessive caution around computers, avoidance, negative feedback about computers, and attempts to minimize computer usage are all indications of anxiety.

Rahman *et al*. conducted a study among academic and non-academic staff to measure the level of their stress [4]. They aimed to identify the differences in stress levels between academic and non-academic staff and the differences in gender in terms of stress. The result of their study showed a moderate level of stress among the respondents, even though the study was limited to one University and one non-academic organization.

Riedi *et al*., in their paper "Technostress about Job Satisfaction and Organizational Commitment Among IT Professionals", determined the impact of technostress on job satisfaction and organizational commitment among IT professionals. The Pearson product-moment coefficient of correlation was used for their statistical analysis of the result. Their result revealed that technostress was negatively correlated to organizational commitment and job satisfaction [5].

Lee *et al*. proposed an empirical study on Technostress among Indian academics. Their work analyzed technostress among Indian academics. This research was conducted on 116 academics in India using an online questionnaire. The study concluded that technostress has significant effects on gender, age, technostress awareness, and tenure of academics [6].

Liu *et al*. proposed an integrated model for investigating the rate at which individuals can be medically affected by technological stress. Their result provided valuable insight and significant knowledge for technostress in healthcare, particularly from academic and practical perspectives [7].

Galluch *et al*. proposed the influence of technostress on the library: a survey of the University for Developmental Studies library. Their work used several variables to focus on the examination of the ICT in library staff, to examine the effect of the use of ICT by library staff. The work aimed at determining and managing technostress among the library staff. The result indicated that the presence of technostress was a negative indication of the overall performance of the library staff [8].

Monica *et al*. proposed a machine learning model for the classification of technostress on Twitter data using predetermined categorical data using SVM, and multinomial naive Bayes baselines models which were finally compared with three Deep Learning models [9].

During the COVID-19 remote work, Derra proposed the Wellbeing Costs of Technology Use: An Investigation using the Italian Translation of the Technostress Creators Scale. Their study shows that the use of remote working increased during the pandemic and invariably increased the risk of technostress among workers [10].

Norhisham proposed an evaluation based on a validated questionnaire. Their work pointed out that occupational risks linked to technostress were high during the pandemic period [11].

Tarafdar *et al.* discussed technostress conceptually, including different dimensions (creators and prohibitions) and how each dimension affects individuals as ICT users. The goal of the research was to raise awareness among employees and employers about the importance of support systems in reducing technostress [12].

However, the existing model as proposed by Liu *et al*. titled Understanding Technostress During the Era of Covid-19: A Conceptual Paper; elaborated on the impact of Digitalization and Adoption of Technology and ICT by Different Employees and the Aftermath on employers. Their work portrayed that employers who already coped with technological usage are seen to be less stressed. Their research focused on the conceptualization of technostress, which included multiple dimensions and how each dimension affected the individuals [13]. The work was resourceful within the developmental framework but was inadequate because of the following reasons:

(i) It was only based on the conceptual view of technostress.

(ii) No dataset was gathered for the evaluation of technostress

(iii) Only a theoretical review was carried out on the evaluations of technostress.

(iv) Characterization and classification of technostress was not carried out in the existing system.

(v) No intelligence approach such as machine learning was adopted in technostress evaluation in the post-COVID era in the existing system.

2. Methodology

This research work adopts a Quantitative methodology using a Random Forest algorithm to evaluate and classification of the impact of techno-stress during and after the COVID-19 era. This study used a survey research approach

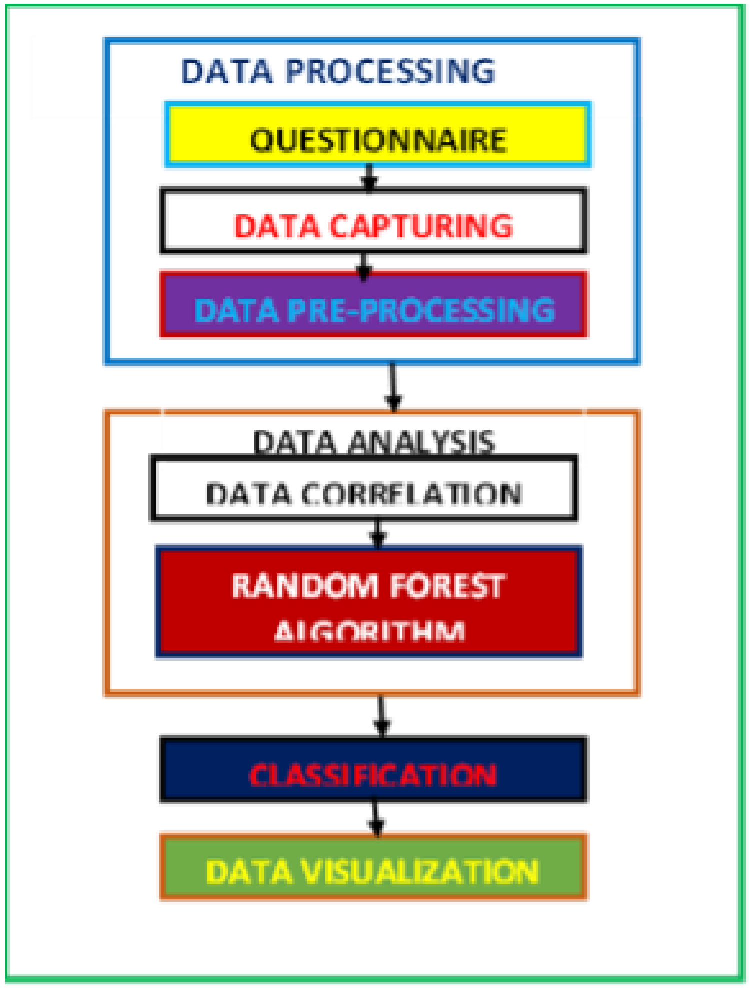

Figure 1: Conceptual framework of the proposed model.

to develop a useful questionnaire for assessing technostress. Random forest (RF) is used to improve the efficiency of technostress assessment. This is because it is a straightforward approach that is based on practical knowledge of the model to be designed, which has advantages over the previous approach, which did not use any analysis model to evaluate technostress that should be used using performance measuring and modeling. Random forest (RF) also increased the performance of the machine learning algorithm by removing associated variables that don't play a role in decision-making [14, 15]. As in regression and classification, the RF algorithm is capable of handling data sets with both continuous and categorical variables. It creates superior fallouts for classification difficulties, which improves visualization. Figure 1 shows the proposed model.

The Model-Based Design (MBD) is a quantitative and conceptual approach to solving challenges in complex control, image analysis, and telecommunication system design. Model formulation is the process of converting our understanding of a natural system into mathematical form [16, 17]. It consists of two parts: the creation of a conceptual model and the transformation of that model into mathematical formulae [18, 19]. Figure 1 shows the framework for the proposed model.

The proposed Random Forest (RF)-based technostress classification framework is structured to ensure accurate stress level prediction and visualization. Below is a detailed breakdown of its practical implementation, including the data collection process, system architecture, algorithm deployment, and real-world applications. The framework begins with a structured data collection process, ensuring high-quality inputs for classification. Information on age, gender, technological experience, and hours spent on technology daily were collected from participants, and self-reported stress levels were captured and categorized into low, moderate, or high stress. To handle missing data, the mean/median imputation for numerical fields and mode imputation for categorical variables (e.g., gender, experience level) were used. Feature engineering was performed on the dataset by converting categorical features (e.g., gender, experience level) into numerical representations using one-hot encoding as well as normalizing numerical values like hours spent on technology.

However, the framework follows a modular pipeline for data ingestion, processing, classification, and visualization. Input validation was done to ensure consistent and complete responses, as well as standardization, feature selection, and cleaning. The training and testing split were 80% for training and 20% for validation. Random Forest (RF) is chosen due to its robustness in handling categorical and numerical data, and this RF was trained on annotated technostress datasets. In the classification results display, stress levels are presented using dashboard visualizations where users also receive personalized feedback on their technostress level. With the availability of the interactive user interface, users can input their parameters and receive stress level predictions in real time.

The description of the Key components of the proposed system are:

2.1. Data capturing/collection

The practice of acquiring data to be examined and utilized for certain reasons later is known as data capture. Particle accelerating, remote surveillance, and simulation software models are examples of high-tech data-collecting approaches (e.g., field-based paper instruments) [20, 21]. It's simpler to share, publish, and cite data that has appropriate information attached at the moment of acquisition. Data capture, also known as electronic data capture, is the process of extracting information from a document and converting it into data that can be connected to a computer [22]. In a broader sense, data capture might refer to obtaining vital information in digital or physical documents.

In this paper, we have segmented our data collection into two sections. The first is through the questionnaire to carry out a proper analysis of technostress, based on the impact or stress that is incurred while using technology during the COVID-19 and post-COVID era. A non-probabilistic sample was utilized to gather a sample of respondents who were capable of answering the questionnaire and were informed about the topic. Even though it might be viewed as biased, this was required for the convenience sampling goal at hand. In an academic setting, carefully structured questionnaires were used to obtain primary data from a variety of people. The sample consists of individuals who are students, lecturers, and other academic institutions that carry and use technology in their different endeavours, especially the academic sector. Some belonged to different institutes and associations relating to the use of technology. The particulars of the respondents obtained from the questionnaires are presented in Table 1. The table represents the demographic variables and technology usage patterns of participants. The following steps were taken to transform this data into structured tables: The data were categorized into different sections to ensure clarity and easy interpretation, including the Demographic Information (Gender, Profession, Age, Education) and the Technology Usage (Frequently Used Technology, Hours Spent on Technology). Each category was transformed into a structured table, making the data easier to read and analyse. The raw data was initially in an unstructured format, combining variables, frequencies, and percentages in a textual layout. To enhance clarity, the data was categorized into Table 1 and Table 2 to improve readability and accessibility. The table is structured with clear headers, values, and labels, making it easier to analyze trends, compare distributions, and perform statistical analysis.

The dataset was transformed into structured tables by selecting features based on relevance, interpretability, and their ability to provide meaningful insights into demographic characteristics and technology usage patterns. Features such as gender, age, profession, education, frequently used technology, and hours spent on technology were chosen for their relevance to the study's objectives, data availability, and variability across responses. Assumptions made during data processing included the accuracy of self-reported information, the distinctiveness of technology usage categories, the representativeness of frequency percentages, and the correct reporting of time spent on technology. At the cause of the data analysis, certain variables were also transformed for better organization, such as consolidating employment status into broader categories and structuring technology usage into defined time intervals. Through these selections and assumptions, the authors were assured of analytical usefulness, minimized redundancy, and improved data interpretability for clearer insights into user demographics and technology engagement.

Table 1 depicts a total of 438 questionnaires collected. Table 1 shows that employees constituted 46.8% of the population sample, 12.1% of the respondents were unemployed and students constituted 41.1% of the respondents respectively. Table 2 shows the various technologies used and the corresponding hours spent on those technology, ranges from 1-3 hours, 5-7 hours, 7-10 hours, and 10-30 hours respectively.

Table 1: The particulars of the respondents obtained from the questionnaires.

Demographic variables	Frequencies	Percentages
Gender		
Male	222	50.7
Female	216	49.3
Profession		
Employee	205	46.8
Unemployee	53	12.1
Students	180	41.1
Age		
18–28 yrs	153	34.9
29–39 yrs	117	26.7
40–50 yrs	92	21.0
51–60 yrs	47	10.7
Education		
FLSC	4	0.9
SSCE	60	13.7
OND	58	9.9
BSC	231	52.7
Master	32	7.3
Others	53	15.5
Frequently used Technology		
Mobile phone	407	30.3
Computer	286	21.3
Security devices	42	33.1
Static equipment	120	8.9
Dynamic equipment	153	11.4
Tele devices	310	23.0
Other gadgets	27	2.0

Table 2: The particulars of the respondents based on hours spent on technology.

Technology	Hours used (in hours)			
	1–3	5–7	7–10	10–30
Mobile phone	1	5	4	10
Computer	18	18	81	27
Security devices	16	14	40	10
Static equipment	13	27	23	3
Dynamic equipment	13	20	23	19
Tele devices	11	11	44	15
Other gadgets	3	11	5	16

2.2. Data pre-processing

Data pre-processing techniques are used for refinement, structuring of data, and feature extraction in an acceptable format for use in machine learning logarithms. The necessity of data pre-processing cannot be overemphasized as its purpose is to validate data for use in machine and statistical learning operations. Most statistical and machine-learning models require the use of numerical data for accurate model definition, prediction, and classification. Contrary to this requirement, the dataset mostly contains features or variables with categorical values. This type of variable is

Table 3: Features of project dataset and its categories.

S/N	Attributes	Description	Data type
1	Gender	Male and female	Nominal
2	Age	18-69	Ordinal
3	Hours spent	1 hour-12 hours	Nominal
4	Technology stressed	Experience with Technology	Nominal
5	Technostress	Stressed or not stressed	Nominal

unsuitable and can render machine learning algorithms inaccurate and unreliable; this shortcoming is overcome by the use of a data pre-processing technique suitable for converting categorical variables to numeric. Notably, for algorithms that do not specifically require the use of numeric data, there are some other forms of preprocessing of the data needed before it is passed on to the training and testing of machine learning models. A sample of raw field data obtained by the researcher is presented in Table 3. It can be observed from Table 3, that there is a need for pre-processing of the dataset; the following techniques used in this research to get the data ready and suitable for model deployment are outlined in the section below:

2.2.1. Data cleaning and filtering

Cleaning and filtering of raw data is an essential step used for discarding unwanted, irrelevant, or incomplete observations in a dataset. Missing values, particularly in large datasets, occur frequently, and it is usually difficult, if not impossible, to identify individual records with missing or incomplete data. Using the data filtering technique, data rows with missing or incomplete entries were detected, and missing values imputation was carried out on the dataset. The cases of null data were also addressed.

2.2.2. Variable transformation

As noted earlier, machine learning models require numeric data types for training, testing, and prediction. In the research dataset, one feature variable (Gender) is described as a categorical variable, therefore, it is required that some form of variable transformation or encoding be done on the data. In this work, a dummy variable transformation method was used. The dummy variable technique converts a categorical feature variable n–1 binary variable where n is the number of classes belonging to each predictor variable; therefore, one dummy variable is created.

2.2.3. Data normalization

Data normalization, also referred to as standardization of data, is a pre-processing technique applied to variables with different scales. This technique ensures that all predictor variables have the same effect on the model outcome.

2.2.4. Training and testing

The project dataset provided the training and testing data. The validation set technique is a frequent method in which the dataset is randomly divided into two sections, usually in a 90:10 split ratio. This split was randomly implemented in Python programming language. The main purpose of splitting the dataset into training and testing data is to create separate unseen data different from the training data that was used during the model training and fitting. The method allows for accurate and proper model evaluation. Other methods for separating data into training and testing include leave-one-out cross-validation, which includes removing one observation from the training set every time, and the K-fold validation procedure, which involves dividing the data into K sets and utilizing K-I for training. Since this study is focused on the prediction and classification of technostress in the COVID and post-COVID era. It's vital to stress that the prediction is for unseen predictor variables rather than those whose data has already been used for training. This is why the dataset is divided into two parts: training data, which will be used for training, and testing data for testing the model fidelity.

2.3. Correlation

Correlation is widely employed in the realm of real-valued sequences. It is applied in various data fields, including real, nominal, and ordinal data. Because it is based on the notion of assigning nominal values to real values, the approach is known as A-correlation (A for assignment).

Table 4: Percentage of the frequently used technology (x) and percentage of hours spent on technology used (y).

Technology	Frequently used technology (%)	Hours spent on technology (%)
Mobile phone	30.3	10
Computer	21.3	27
Security devices	33.1	10
Static equipment	8.9	3
Dynamic equipment	11.4	19
Tele devices	23	15
Other gadgets	2	16

The statistical measure of the relationship between two variables is called correlation [23]. This metric works well with variables that exhibit a linear relationship with each other. The correlation coefficient quantifies the strength of the relationship between two variables. This can be calculated using the following formula:

$$\tau_{xy} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 (y_i - \overline{y})^2}}, \tag{1}$$

where:
τ_{xy} – the correlation coefficient of the linear relationship between the variables x and y,
x_i-the values of the x-variable in a sample,
$\overline{x}$- the mean of the values of the x-variable,
$y_i -$ TThe values of the y-variable in a sample,
$\overline{y}$– the mean of the values of the y-variable.

To determine how frequently used technology correlates with hours used in other to determine the level of stress devices can exert on users, the following steps are necessary.

Step 1. To obtain a data sample with the values of the x-variable and y-variable, the data sample in Table 4 was used.

Step 2. The means (averages) $\underline{x}$ for the x-variable and $\underline{y}F$ for the y-variable were calculated respectively;

$$\overline{x} = \frac{x^1 + x^2 + x^3 + x^4 + x^5 \pm \cdots + x^n}{n}. \tag{2}$$

Hence, the mean average for technology used = 18.6

$$\overline{y} = \frac{y^1 + y^2 + y^3 + y^4 + y^5 \pm \cdots + y^n}{n}. \tag{3}$$

Also, the mean average for hours spent = 14.3

Step 3. For the x-variable, the mean from each value of the x-variable (that is the new variable “e”) was subtracted. The same was repeated for the y-variable (that is the variable “f”).

Step 4. Each e-value was multiplied by the corresponding f-value, and the sum of these multiplications (the final value is the numerator in the formula) was evaluated.

Step 5. Each e-value was squared, and the sum of the result was calculated.

Steps 3 to 5 are depicted in the correlation table (Table 5).

Finally, using the obtained numbers, the correlation coefficient was calculated by applying the formula in eqn. (1) as follows:

$$r_{xy} = \frac{-33.9}{\sqrt{795.3} \times 351.4} = 0.00012. \tag{4}$$

Therefore, the coefficient indicates that the most used technology and hours spent do not have a high positive correlation. This means that the frequently used technology does not have a greater impact or a determining factor on the hours spent.

Table 5: Correlation analysis.

Technology	X	Y	E	F	E*f	e^2	f^2
Mobile phone	30.3	10	11.7	-4.3	-50.31	136.89	18.49
Computer	21.3	27	2.7	12.7	34.29	7.29	161.29
Security Devices	33.1	10	14.5	-4.3	-62.35	210.25	18.49
Static Equipment	8.9	3	-9.7	-11.3	109.61	94.09	127.69
Dynamic Equipment	11.4	19	-7.2	4.7	-33.84	51.84	22.09
Tele Devices	23	15	-4.4	0.7	-3.08	19.36	0.49
Other gadgets	2	16	-16.6	1.7	-28.22	275.56	2.89

2.4. *Random forest algorithm*

A supervised learning model called the random forest algorithm was used to classify unlabeled data by labeling it. Regression and classification issues are resolved with the usage of the random forest technique [24–26].

Random forest models were used for the classification of the impact of technostress in the COVID-19 and post-COVID-19 era. After correlation analysis of the dataset, that is, a measure of correlation between the predictor variables and the explanatory was carried out to completely and accurately assess the impact of each feature variable on the dependent variable.

The Random Forest model equation is described as $y = a_0 + a_1 x_1 + c$ The estimated values for the dependent variable are given as technostress = a_o + a_1 x age + c. a_o and a_1 are the intercepts and coefficient of the predictor variable respectively. The ordinary least square method is used for fitting and optimizing the regression line by minimizing the residual error. The residual error is the square of the difference between the actual and the predicted values of the independent variable. The estimated values of the impact of technostress are given as:

$$\text{Technostress} = \widehat{\alpha}_0 + \widehat{\alpha}_1 x \text{ Gender} + \widehat{\alpha}_2 x \text{ age} + \widehat{\alpha}_3 x \text{ hourspent} + \widehat{\alpha}_4 x \text{ tech used} + \hat{c}. \quad (5)$$

The notation $\alpha_0 \rightarrow \alpha_4$ represents the estimated slopes for all the predictor variables, while the error is expressed as $c_i = y_i - \hat{y}$. The data was fitted to the model using the ordinary least square (OLS) approach. The sum of the squared residuals is minimized using the OLS technique. The regression line, commonly known as the residual sum of squares (RSS), is the estimate that minimizes the sum of squared residual values. It is written as follows:

$$\sum_{i}^{n} = 1(y_i - \hat{y}_i)^2. \quad (6)$$

$$\sum_{i}^{n} = 1(y_i - \hat{a}_o + \hat{a}_1 x \text{ Gender} + \hat{a}_2 x \text{ age} + \hat{a}_3 x \text{ hourspent} + \hat{a}_4 x \text{ techused})^2. \quad (7)$$

The estimated coefficients $\hat{a}_o \rightarrow \hat{a}_4$ are coefficients used for the minimization of the RSS. The coefficients are estimated using the ordinary least square method, the same as with the Random Forest Model.

2.4.1. *Assessing the linear regression model*

The root mean squared error (RMSE) approach was used to evaluate the RF model. The square root of the average error is predicted as $\hat{y}_i$ values is RSME, and is expressed as

$$RSME = \sqrt{\frac{\sum_i^n = 1\ (y_i - \hat{y}_i)^2}{n}}. \quad (8)$$

The total number of observations is n; this metric assesses the model's overall accuracy and serves as a benchmark for comparing other models. Similarly, the residual standard error (RSE) may be used to evaluate the model's accuracy. The fundamental distance between RMSE and RSE is that the degrees of freedom are considered rather than the amount of data. RSE is expressed in the equation below:

$$RSE = \sqrt{\frac{\sum_i^n = 1\ (y_i - \hat{y}_i)^2}{(n - p - 1)}}. \quad (9)$$

The entire number of observations (result) is n, while the total of predictor variables is p. By computing the difference between the actual and predicted values, the mean absolute error (MAE) may be used to evaluate the effectiveness of a regression model. MAE may be expressed as:

$$MAE = \sqrt{\frac{\sum_i^n = 1\ (y_i - \hat{y}_i)^2}{n}}. \tag{10}$$

In addition, R-squared statistics was utilized to evaluate the linear regression model. The R^s Assess the coefficient of fit of the model to the provided data, which is expressed in percentage. This metric calculates the technostress proportionate variance to the predictor variables. It is mainly used to explain how well a linear regression model fits the data.

$$R^2 : R^2 = 1 - \frac{\sum_i^n = 1(y_i - \hat{y}_i)^2}{\sum_i^n = 1(y_i - \hat{y})^2}. \tag{11}$$

2.4.2. Correlation analysis

A measure of correlativity showed that it was the only predictor variable with statistical significance on the explanatory variable. The measure of correlation was done using the Pearson correlation formula expressed as follows:

$$P = \frac{\sum h - m_h(\text{technostress} - \ m_{\text{technostress})}}{\sqrt{\sum (h - m_h)^2 \sum (\text{technostress} - m_{\text{technostress}})^2}}. \tag{12}$$

The significance level of the correlation (P-value) was determined using the degrees of freedom from the correlation coefficient table and/or calculating the t-value.

2.5. Classification algorithm

Machine learning can be classified as supervised, unsupervised, and reinforced learning. A classification algorithm is a supervised learning technique that is used to categorize newly trained data. In the classification algorithm, a discrete output function (y) is mapped to input variables (x). types of classification: binary classifier and multi-class classifier. Learners in classification problems and lazy and eager learners. Types of ML classification algorithms: ML classifiers are divided into two categories: linear models and non-linear models. The linear model consists of logistics regression and support vector machines. The nonlinear model consists of K nearest neighbours, kernel SVM, naive Bayes, decision tree classification, and random forest classification. At the cause of the classifications, the following steps were considered:

Step 1. Ensure the collection of measures, an instruction sequence, and an authentication sequence.

- i. The bootstrap form to select a variable at random from the available sample was used while adhering to predetermined standards.
- ii. The datasets were observed to note each dataset that appeared to have an efficiency of n or roughly the same as the original dataset.
- iii. The remaining set was used as a testing dataset, and K sets were deployed as samples for training datasets for each of the variables.

Step 2. Beginning the development of the RF classifier.

- i. A model was created by feeding in K training samples and applying the RF algorithm;
- ii. A combination of categorization approaches was developed by employing the decision tree for categorization K.

Step 3. Use the feedback from the assessment package to gauge the model's efficiency.

Step 4. Finally, the unmarked sampled items were chosen by sorting every observation using only the random forest in the direction of the F-measure. A weighted vote within the classification performance of each sub-classifier was used to determine the final result.

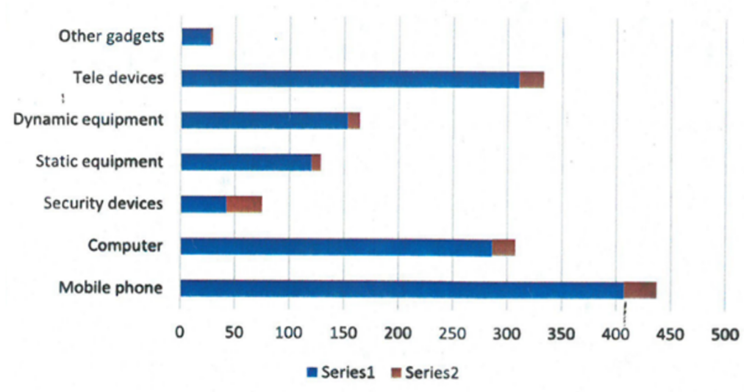

Figure 2: Technology affected usage level.

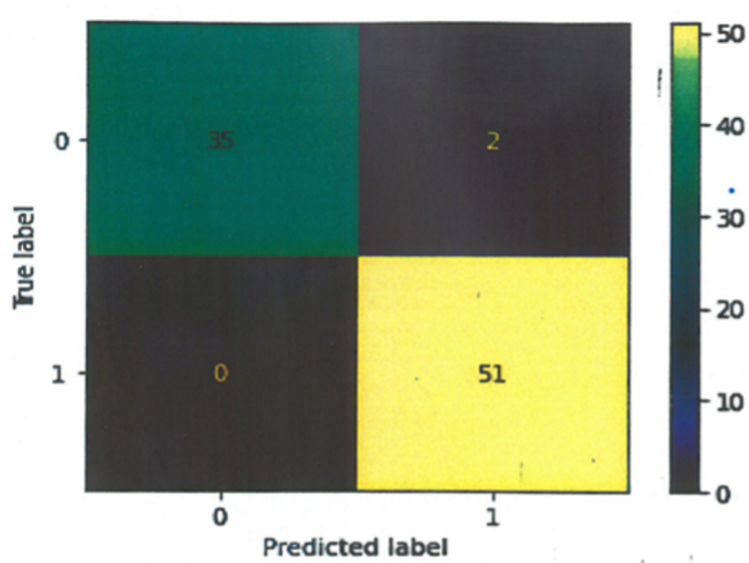

Figure 3: Graph plot of the predicted value and true value.

2.6. *Data visualization*

Information perception is the graphical representation of data and information. By applying observable signals such as outlines, diagrams, and guides, data visualization tools make it simple to examine and comprehend patterns, expectations, and instances in data. In the Big Data era, information perception techniques and assets are critical for reviewing massive amounts of data and making data-driven decisions and suggestions. Figure 2 shows the technology-affected usage level, Figure 3 shows the graph plot of the predicted value and true value, Figure 4 shows the graph plot of the predicted value and true value, and Figure 5 shows the Plotting of the ROC curve and trained data.

Figure 2 illustrates the level of technology usage across various device categories. It categorizes usage levels for different types of devices, including mobile phones, computers, security devices, static equipment, dynamic equipment, telecommunication (tele) devices, and other gadgets. Reveals that mobile phones and computers are the most frequently used technologies, with mobile phones leading by a significant margin. The disparity in usage levels between Series 1 and Series 2 for each device suggests different usage patterns, potentially reflecting shifts in technology reliance due to changing circumstances like the transition from COVID-19 to post-COVID-19 environments.

In Figure 3, the model's high count of true positives and true negatives (35 + 51) versus low false predictions (only 2 false positives and 0 false negatives) indicates strong overall accuracy. With only 2 false positives, the model has a high precision for the positive class. The absence of false negatives indicates a recall of 100% for the positive class, meaning the model reliably detects all positive instances. This confusion matrix suggests that the model is effective in

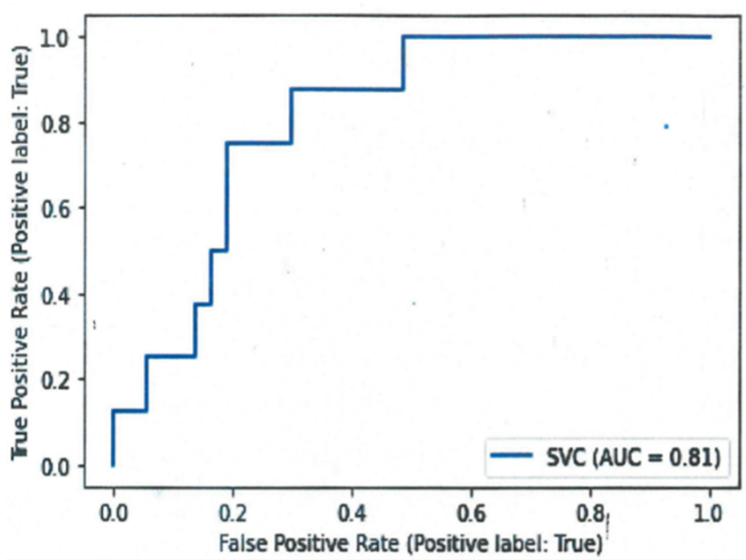

Figure 4: Plotting of ROC curve.

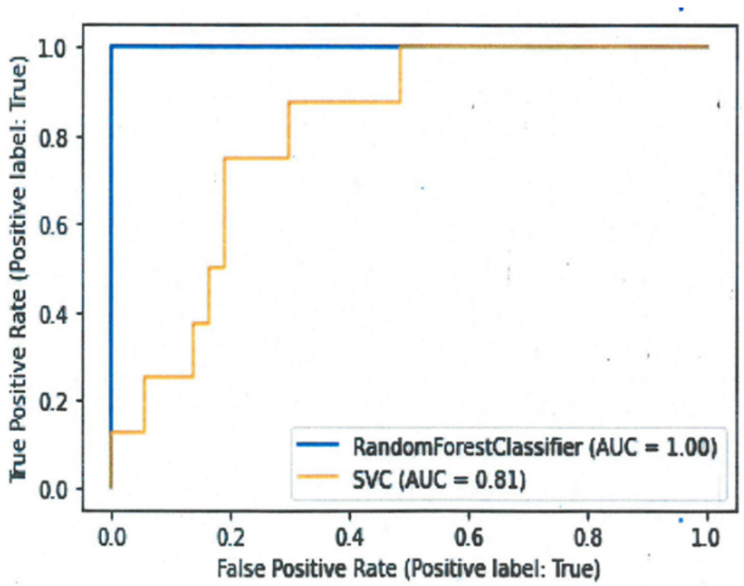

Figure 5: Plotting of ROC curve and trained data.

distinguishing between the two classes with minimal errors, making it a reliable tool for binary classification in this context.

Figure 4 shows a high concentration along the diagonal, indicating that the model is performing well.

An AUC of 1.0 indicates a perfect model, while an AUC of 0.5 suggests no better than random guessing. The Random Forest's AUC of 1.00 indicates that it is an excellent classifier on this dataset. The SVC's AUC of 0.81 still shows decent classification ability but indicates that the RF model is more accurate for this specific task.

2.7. Model implementation

Both linear and K-nearest neighbour models were used to explain the relationship between the predictor(s) and the outcome variables (technostress). The goal was to understand the general relationship between the variables rather than the model's predictability. Based on the summary of the linear regression and KNN model, the following conclusions were reached:

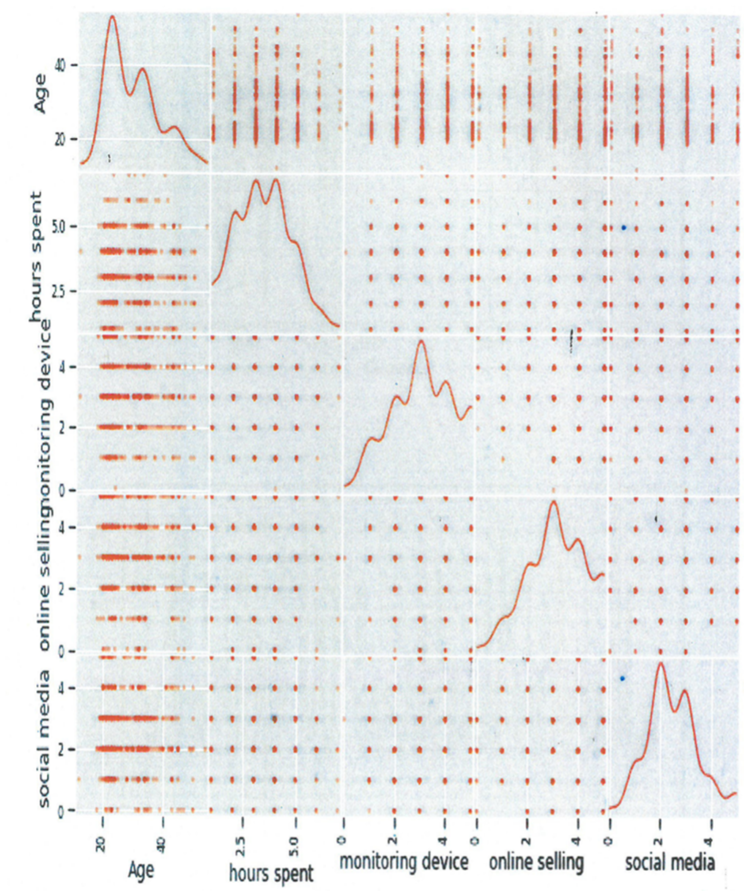

Figure 6: Joint plotting of the dataset.

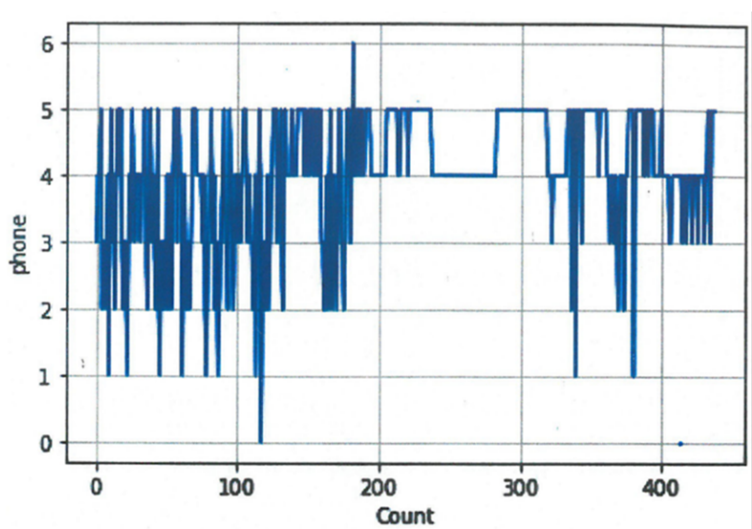

Figure 7: Line graph of phone and count.

1. Intercept: The intercept α_0 for Random Forest models was obtained to be 0.606. In the model, the intercept had a p-value above 0.05, implying that there was no statistical significance on the predicted outcome (Technostress).
2. Analysis of variance: Models had an F-statistics score that is greater than one, and also, the p-value obtained in

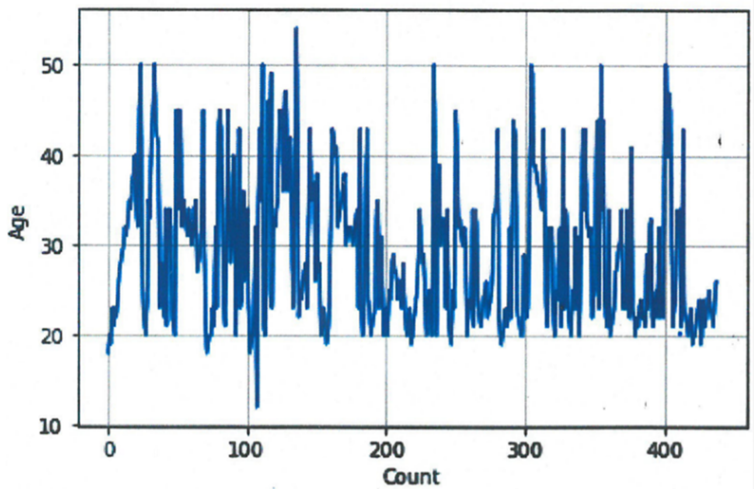

Figure 8: Line graph of age and count.

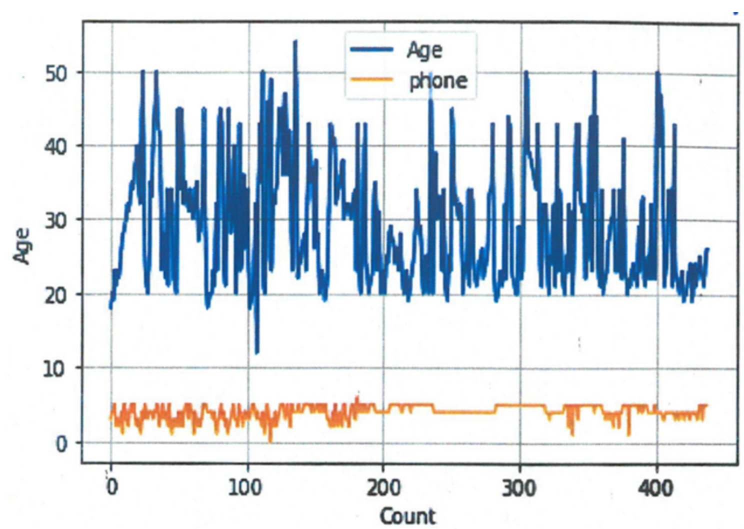

Figure 9: Line graph of age and phone.

models was below the threshold value of 0.05.

3. Analysis of impact of predictor variables: In this model, only 'technostress' had a statistical significance. In other words, out of the four predictor variables (gender, age, hours spent, and experiences using technology), experiences using technology (technostress) are statistically significant.

3. Result and discussions

The accuracy of the regression models was evaluated using four different methods: Mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and R squared (R^2)

(i) Mean Absolute Error (MAE): Absolute error is a straightforward statistics metric for determining the difference between true and predicted values. The mean absolute error was calculated for the full dataset using the average. The MAE is more resilient to outliers since it uses the same unit as the output variables.

(ii) Mean Square Error (MSE): The MSE calculates the square error between actual and predicted values. The goal of MSE is to keep negative values from being canceled during evaluation. MSE can be used as a loss function, but the result is the square of the output and is not robust against outliers.

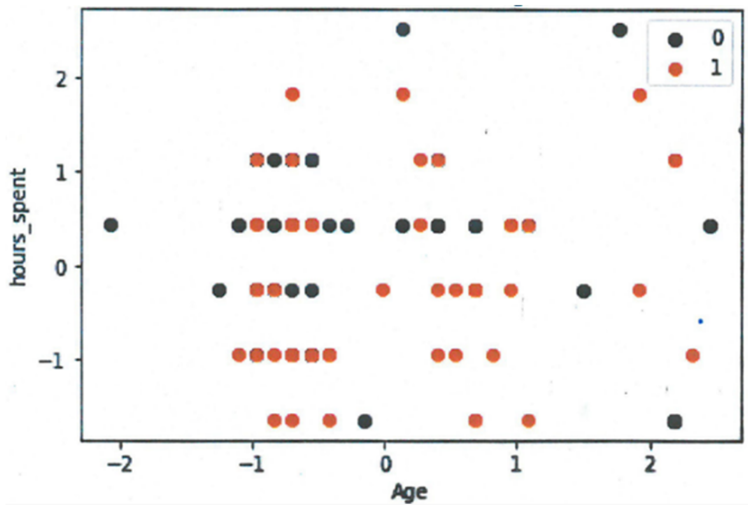

Figure 10: Line graph of hours-spent and age.

Table 6: Model types and accuracy scores.

Model Type	MAE	MSE	RMSE	R^2	RSS
Random Forest Algorithm	0.0455	0.0310	0.1761	0.001	25.4

(iii) Root Mean Square Error (RMSE): The square root of MSE is RMSE, as the name implies. Although it is expressed in the same unit as the output variable, it is not robust to dataset outliers.

(iv) R-Squared R^2: R^2 as mentioned is a metric for determining how well data fits a model. Unlike the preceding methods, R^2 models can be compared. R^2 is also known as the coefficient of determination or goodness of fit, and it calculates how much better a regression line is than a mean line. Table 6 presents a summary result of the measurement of accuracy for the Random Forest algorithm.

Additionally, the RF algorithm was used in the aspect of classifying individual gender (i.e., Male or Female) based on Technostress, age, hours spent, and experiences with technology.

In using RF for training data, the input data was the feature of questionnaires, and the classes were the technologies used. The class of the test data, which contains input variables, was determined by comparing the distance of the known tuple to the random forest of that reference model. Figure 6 shows that there are some relationships between the variables, particularly between social media usage, online selling, and age. There is a slight negative correlation between age and hours spent on social media and online selling. This shows that as age increases, individuals tend to spend less time on these activities. The distribution of hours spent on various activities is also right-skewed, indicating that most individuals spend a relatively short amount of time on these activities. The distribution of monitoring device usage is less clear, with no apparent patterns or correlations with other variables. This is due to various factors, such as the type of monitoring device used or individual preferences.

The graph in Figure 7 exhibits a highly fluctuating pattern with frequent ups and downs. This shows that the frequency of phone-related events or activities is not consistent over time.

Figure 8 appears to be a line graph illustrating the relationship between an ”Age” variable and a ”Count” variable. The graph shows a fluctuating pattern over time, with the ”Age” variable increasing and decreasing in a somewhat cyclical manner. The graph exhibits a highly fluctuating pattern, with frequent ups and downs. This shows that the ”Age” variable is not consistently increasing or decreasing over time.

In Figure 9, the ”Age” line shows a fluctuating pattern, with periods of increase and decrease. There is no clear upward or downward trend, suggesting that the age of the subjects or data points is not consistently increasing or decreasing over time.

In Figure 10, the Random Forest model is not overfitted to the training data, this means it generalizes well to new, unseen data.

4. Conclusion

This paper focused on analyzing quantitative data and pre-processing techniques to develop an enhanced machine learning model for characterizing and predicting technostress in the COVID and post-COVID era. The Random Forest (RF) model was utilized to predict technostress across various scenarios, achieving the highest mean performance estimates and the lowest uncertainty estimates. Emphasis was placed on accurately classifying individuals with low or high stress levels. Additionally, the RF model was applied to classify gender based on the dataset, achieving an accuracy of 90%, and demonstrating its efficiency in the classification process. This study successfully developed a model for effective and efficient analysis of technostress, particularly in the educational sector, where remote presence methodologies were adopted. Given the proven impact of technology on workplace stress, it is crucial for organizations that have integrated digital processes to conduct regular assessments of their employees to mitigate the effects of technostress.

In real-world scenarios, our method proves its effectiveness in deriving meaningful insights, strengthening its applicability across industries. For instance, in the financial industry, customer segmentation based on demographic and technology usage patterns can improve fraud detection and enhance personalized banking services. A study on retail businesses could analyze consumer behaviour by tracking mobile and computer usage trends, leading to more effective marketing strategies and customer engagement. In the education sector, researchers could examine how students' reliance on mobile phones and computers affects learning efficiency, allowing institutions to develop targeted digital learning tools. Additionally, healthcare facilities can assess the impact of security devices and telemedicine usage among different demographic groups to optimize remote patient care services.

Future research should explore the application of diverse and hybrid machine learning models to assess the impact of technology usage in other organizational contexts beyond education.

Data availability

The data used in this study is available upon reasonable request. Due to privacy and confidentiality considerations, access to the dataset may be restricted. However, researchers who wish to replicate or extend this work may contact the corresponding author for further details regarding data access and usage conditions.

References

[1] G. G. James, J. E. Asuquo, N. A. Michael & E. O. Etim, "Adaptive predictive model for post Covid'19 health-care assistive medication adherence system", in *Contemporary Discourse on Nigeria's Economic Profile*, Ayandel *et al*. (Ed.), A Festschrift in Honour of Prof. Ukpabio Ndaeyo Nyaudoh, A Publication of Universit of Uyo, Uyo. Akwa Ibom State, Nigeria, 2023, pp. 622-631. https://www.researchgate.net/profile/James-G/publication/380941779.

[2] G. James, I. Umoren, A. Ekong, S. Inyang & O. Aloysius, "Analysis of support vector machine and random forest models for classification of the impact of technostress in COVID and post-covid era", Journal of the Nigerian Society of Physical Sciences **6** (2024) 2102. https://doi.org/10.46481/jnsps.2024.2102.

[3] C. T. Cuervo, M. N. Orviz, G. S. Aree & S. I. Femandez, "Tecnostres en la sociedad de la tecnologya y la comunicacion: revision bibliografica a partir de la web of science", Archivos de Prevencion de Riesgos Laborales **21** (2018) 25. https://dx.doi.org/10.12961/aprl.2018.21.01.4.

[4] M. M. Rahman, F. Khatun, A. Uzzaman, S. S. Sami, M. A. Bhuiyan & T. S. Kiong, "A comprehensive study of artificial intelligence and machine learning approaches in confronting the coronavirus (COVID-19) pandemic", International Journals of Health Services **51** (2021) 446. https://doi.org/10.1177/00207314211017469.

[5] R. Riedi, H. Kindermann, A. Auinger & A. Javor, "Technostress from a neurobiological perspective - system breakdown increases the stress hormone cortisol in computer users", Business & Information Systems Engineering **4** (2012) 61. https://aisel.aisnet.org/bise/vol4/iss2/2.

[6] Y. K. Lee, C. T. Chang, Z. H. Cheng & Y. Lin, "Helpful-stressful cycle? Psychological links between type of mobile phone user and stress", Behaviour & Information Technology **35** (2016) 1055800. https://doi.org/10.1080/0144929X.2015.1055800.

[7] C. Liu, T. Cheng & C. Chen, "Exploring the factors that influence physician technostress from using mobile electronic medical records", Informatics for Health and Social Care **44** (2017) 1364250. https://doi.org/10.1080/17538157.2017.1364250.

[8] P. Galluch, V. Grover & J. Thatcher, "Interrupting the workplace: examining stressors in an Information technology context", Journal of the Association for Information Systems **16** (2015) 47. https://open.clemson.edu/all_dissertations/448.

[9] M. Monica, E. Ingusci, F. Signore, A. Manuti, M. L. Giancaspro, V. Russo, M. Zito & C. G. Cortese, "Wellbeing costs of technology use during covid'19 remote working: an investigation using the Italian translation of the technostress creators scale", Sustainability **12** (2020) 5911. https://doi.org/10.3390/su12155911.

[10] N. D. Derra, *Machine learning and technostress as important aspects for improving the performance of data scientists in contemporary marketing contexts*, Ph.D. thesis, Business and Economics, Faculty of Law, University of Bayreuth, 2021. https://doi.org/10.15495/EPub_UBT_00005398.

[11] N. Norhisham, "Understanding technostress during the era of Covid'19: a conceptual paper", International Journal of Academic Research in Business and Social Sciences **11** (2021) 1936. https://doi.org/10.6007/IJARBSS/v11-i8/10628.

[12] M. Tarafdar, C. I. Cooper & J. F. Stich, "The technostress trifecta – techno eustress, techno distress, and design: theoretical directions and agenda for research", Information Systems Journal **29** (2019) 12169. https://doi.org/10.1111/isj.12169.

[13] C. Liu, T. Cheng & C. Chen, "Exploring the factors that influence physician technostress from using mobile electronic medical records", Informatics for Health and Social Care **44** (2017) 1364250. https://doi.org/10.1080/17538157.2017.1364250.

[14] G. James, A. Ekong, E. Abraham, E. Oduobuk & P. Okafor, "Analysis of support vector machine and random forest models for predicting the scalability of a broadband network", Journal of the Nigerian Society of Physical Sciences **6** (2024) 2093. https://doi.org/10.46481/jnsps.2024.2093.

[15] C. Ituma, G. G. James & F. U. Onu, "Implementation of intelligent document retrieval model using neuro-fuzzy technology", International Journal of Engineering Applied Sciences and Technology **4** (2020) 65. http://dx.doi.org/10.33564/IJEAST.2020.v04i10.013.

[16] G. G. James, E. G. Chukwu, P. O. Ekwe, E. C. Asogwa & C. H. Darlington, "Design of an intelligent-based system for the diagnosis of lung cancer", International Journal of Innovative Science and Research Technology **8** (2023) 796. https://www.research-gate.net/profile/James-G/publication/372338202.

[17] G. James, A. Ekong & H. Odikwa, "Intelligent model for the early detection of breast cancer using fine needle aspiration of breast mass", International Journal of Research and Innovation in Applied Science (IJRIAS) **9** (2024) 348. https://doi.org/10.51584/IJRIAS.2024.90332.

[18] G. G. James, A. Ejaita, C. Ituma & J. E. Asuquo, "Development of hybrid intelligent based information retrieval technique", International Journal of Computer Applications **184** (2022) 1. https://doi.org/10.5120/ijca2022922401.

[19] G. Zhang, F. Zhou, F. Wang & J. Luo, "Knowledge creation in marketing based on data mining", 2008 International Conference on Intelligent Computation Technology and Automation (ICICTA), Changsha, China, 2008, pp. 782-786. https://doi.org/10.1109/ICICTA.2008.45.

[20] Z. L. Ping & S. Q. Liang, "Data mining application in banking-customer relationship management", 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), Taiyuan, China, 2010, pp. 340-342. https://doi.org/10.1109/ICCASM.2010.5619002.

[21] E. W. T. Ngai, L. Xiu & D. C. K. Chau, "Application of data mining techniques in customer relationship management: a literature review and classification", Expert Systems with Applications **36** (2009) 2592. https://doi.org/10.1016/j.eswa.2008.02.021.

[22] A. L. Roth, M. Huff & K. Bendib, *Censored 2018: press freedoms in a "Post-Truth" world*, The Top Censored Stories and Media Analysis of 2016-17, New York: Seven Stories Press, New York, 2017. https://www.lighthousebookshop.com/book/9781609807818.

[23] H. Timonem & J. Vuori, "Visibility of work: how digitalization changes the workplace", Proceedings of the 51st Hawaii International Conference on System Sciences, Waikoloa Village, HI, USA, 2018. http://hicss.hawaii.edu/.

[24] G. G. James, G. P. Oise, E. G. Chukwu, N. A. Michael, W. F. Ekpo & P. E. Okafor, "Optimizing business intelligence system using big data and machine learning", Journal of Information Systems and Informatics **6** (2024) 631. https://doi.org/10.51519/journalisi.v6i2.631.

[25] C. Ituma, G. G. James & F. U. Onu, "A neuro-fuzzy based document tracking & classification system", International Journal of Engineering Applied Sciences and Technology **4** (2020) 414. http://www.ijeast.com.

[26] G. G. James, A. Ekong, E. Abraham, E. Oduobuk, N. Michael, V. Ufford & O. Ebong, "An enhanced control solution for efficient urban waste management using deep learning algorithms", African Scientific Reports **3** (2024) 183. https://doi.org/10.46481/asr.2024.3.3.183.