1. Tasneem SOFRI[1,2,3], 2. Allan Melvin ANDREW[2,4], 3. Hasliza A RAHIM[1,2,5], 4. Hiromitsu NISHIZAKI[3], 5. Latifah Munirah KAMARUDIN[1,6], 6. Peng Wen WONG[7], 7. Ping Jack SOH[8]

Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia (1), Advanced Communication Engineering, Centre of Excellence (ACE), Universiti Malaysia Perlis, Perlis, Malaysia (2), Integrated Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi, Yamanashi, Japan (3), Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia (4), Faculty of Graduate Studies (FGS), Daffodil International University, Dhaka, Bangladesh (5) Centre of Excellence for Advanced Sensor Technology (CEASTech), Universiti Malaysia Perlis, Perlis, Malaysia (6), Filpal (M) Sdn Bhd, Sains@Universiti Sains Malaysia, Penang, Malaysia (7), Centre for Wireless Communications (CWC), University of Oulu, Oulu, Finland (8)
ORCID: 1.0000-0003-0218-2584; 2.0000-0002-0377-0006; 3. 0000-0002-5662-7158; 4. 0000-0002-7717-8312; 5. 0000-0002-2547-3934; 6. 0009-0005-5550-6521; 7. 0000-0002-6394-5330

doi:10.15199/48.2024.06.23

Enhancing Predictive Models for Assessing 5G Exposure Effects on Human Health and Cognition through Supervised Machine Learning: A Multi-Stage Feature Selection Approach

***Abstract**. No prior reviews have focused on any comprehensively examine the effects of 5G exposure (700 MHz to 30 GHz) on human health and cognition using supervised Machine Learning (ML). This novel research combined the Multi-Stage Feature Selection (MSFS) and hybrid features for classification machine learning model. The approach which includes the use of MSFS, yielded better results in terms of accuracy, precision, F1-score, sensitivity, and specificity when contrasted with the approach that did not incorporate MSFS with accuracy more than 0.95 for both datasets.*

***Streszczenie.** Żadne wcześniejsze przeglądy nie skupiały się na kompleksowym badaniu wpływu narażenia na sieć 5G (700 MHz do 30 GHz) na zdrowie ludzkie i funkcje poznawcze przy użyciu nadzorowanego uczenia maszynowego (ML). W tym nowatorskim badaniu połączono wieloetapowy wybór cech (MSFS) i funkcje hybrydowe na potrzeby modelu uczenia maszynowego klasyfikującego. Podejście obejmujące wykorzystanie MSFS dało lepsze wyniki pod względem dokładności, precyzji, współczynnika f1, czułości i specyficzności w porównaniu z podejściem, które nie obejmowało MSFS z dokładnością większą niż 0,95 dla obu zbiorów danych (**Udoskonalanie modeli predykcyjnych do oceny wpływu narażenia na sieć 5G na zdrowie ludzkie i funkcje poznawcze poprzez nadzorowane uczenie maszynowe: wieloetapowe podejście do wyboru funkcji**)*

Keywords: antenna and propagation; bioelectromagnetic; probabilistic neural network; supervised machine learning
Słowa kluczowe: antena i propagacja; bioelektromagnetyczny; probabilistyczna sieć neuronowa; nadzorowane uczenie maszynowe.

Introduction

Technologies for wireless or mobile communication have developed into indispensable tools for daily communication. They enable billions of people throughout the world to keep in touch, travel in and out of cities with safety, and watch free-to-air television in their homes. The Internet is a vital resource in many sectors nowadays. 5G is seen as ushering in yet another new era and went into widespread use since 2019. The projected benefits of 5G include improved e-Health and a wide range of new applications such as telemedicine, remote surveillance, telesurgery, self-driving cars and road safety, smart homes and buildings, smarter and cleaner cities, other intelligent transport systems, 3D video, cloud computing and performance, virtual and augmented reality, and massive machine-to-machine communications for industry automation and manufacturing [1]. Supporting these services applications on 3G and 4G networks is currently challenging.

There has been some public worry about the potential health dangers related to using mobile phones and living close to base stations since the development of mobile communication technologies. Radiofrequency Electromagnetic Radiation (RF-EMR) from frequently used wireless devices, such as cell phones, cordless phones, Wireless Fidelity (Wi-Fi) routers, and cell tower infrastructure has been linked to undesirable health consequences as this debate is still an issue [2]. MP Radiofrequency Radiation (RFR) has been labelled a "Possible Human Carcinogen" (Group 2B) based on comprehensive in vitro, in vivo, and epidemiological studies by the World Health Organization (WHO) and the International Agency for Research on Cancer (IARC) in 2011 [3].

Russell (2018) mentioned from the conclusion of their recent peer-review paper that although 5G technology may have a wide range of applications and advantages, it is also becoming increasingly obvious that if it is widely used, serious adverse effects to human health and ecosystems may happen. Wavelengths of the current radiofrequency radiation to which humans are exposed appear to be toxic to biological systems [2]. Before the additional 5G roll-out, more than 230 scientists from more than 40 nations [4] have already voiced their "severe worries" about the pervasive and growing exposure to Electromagnetic Field (EMF) produced by electric and wireless devices. Numerous recent scientific articles have demonstrated that EMF impacts living beings at levels much below the majority of international and national recommendations, according to them [5].

As highlighted by R. N. Kostoff et al. (2020) that before further rollout can be justified, much more analysis and testing of potential 5G health consequences under actual usage situations is necessary [6]. There is a growing amount of research focused on the application of Machine Learning (ML) in 5G networks [7], [8]. With the increased speed and lower latency of 5G, there is an opportunity to use ML to optimize network performance, improve network security, and enhance user experience. One of the areas of research is focused on using ML to optimize network resource allocation in 5G networks [9]. This includes using techniques such as reinforcement learning to optimize the allocation of network resources in real-time, based on changing network conditions and user demands. Another area of research is focused on using ML to enhance security in 5G networks. This includes using ML algorithms to detect and mitigate network attacks, as well as using ML to identify potential security threats and vulnerabilities in the network [10]. There is also research being conducted on the use of ML to improve the overall user experience in 5G networks [11]. This includes using ML algorithms to predict user behavior and preferences, to provide more personalized services and content. Overall, the combination of ML and 5G has the potential to revolutionize the way that networks are managed and optimized, leading to improved performance, enhanced security, and a more satisfying user experience.

The conventional method to validate whether there is an effect of Radiofrequency-Electromagnetic Field (RF-EMF) on human health is by conducting manual analysis using the statistical technique analyses [12]–[26] through recent peer-reviewed articles. The statistical analyses are performed by comparing the assessed parameters under no exposure and exposure of the RF-EMF signal. The *p*-values are calculated using statistical technique analyses such as Analysis of Variance (ANOVA), independent t-test, Pearson Chi-Square, and Wilcoxon signed rank tests. If there was significant difference between the values of the investigated parameters under no exposure and exposure, the *p*-values are less than 0.05 ($p<0.05$). This indicates that there is an RF-EMF effect on the investigated parameters. If $p>0.05$, this indicates that no RF-EMF effect on the investigated parameters. However, these manual statistical technique analyses led to time consuming and with implementing ML in the bioelectromagnetic research can attempts to discover the undiscovered pattern in data as well as aims to address users to make intelligent judgements from their research outcomes. Furthermore, ML advances the use of prediction tools to support future health checks (ex-vivo) and enables researchers to see how environmental factors may affect a final decision [27].

The exploration and exploitation of the data will be insufficient during the feature selection as the features are reduced at the initial stage. As a result, only some redundant features are selected, and some useful features are lost due to poor data management. The proposed multi-stage approach consists of feature engineering within natural language processing, signal reconstruction, feature selection, feature extraction, improved learning techniques for resampling and cross-validation, and the configuration of hyperparameters.

Previous researchers depicted the use of conventional feature selection method, basically, by using a single-stage feature selection method. In the single-stage feature selection method, the important features are extracted from the raw data, and the extracted data is further filtered to select only important and useful features [28], [29]. By solving the problem in a novel manner, a MSFS with Artificial Intelligence (AI) algorithm is utilized to analyze collected data systematically and make reasonable conclusions, making the whole process automatic. However, to the best of the author's knowledge, none of the previous reviews focused on the health effects resulting from the exposure from the 5G MP and BS antennas from 700 MHz to 30 GHz on the cognitive performance and the human physiological parameters utilizing ML algorithms, especially the supervised learning in the scope of prediction model with result to develop high accuracy classifiers for predicting the potential impact of RF-EMF exposure on human in epidemiological studies.

The main contributions of this paper include the following:

i. The modified MSFS designed depend on database from the assessment data parameter of 5G BS exposure experiment.
ii. The validated performance of the proposed classifier in terms of classification accuracy, precision, F1-score, sensitivity, and specificity.

Methodology

This study of hybridized MSFS framework as shown in Figure 1 and supervised ML for 5G BS health effect detection classification was done to the physiological measurements of the individuals in terms of body temperature, systolic blood pressure, diastolic blood pressure, and heart rate, as well as four cognitive performance outcomes are the existing data from the prior studies used an Electric (E)-field of 1 V/m for 5G signals operating at 700 MHz and 3.5 GHz, and 0.64 V/m for 28 GHz frequencies.

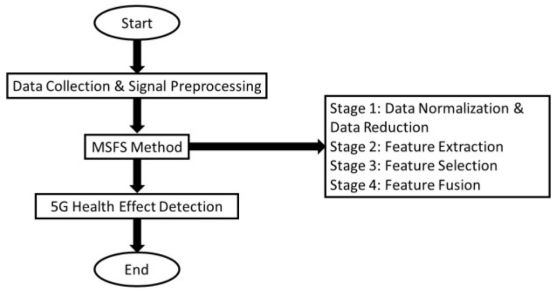

Fig.1. Main framework of 5G BS health effect detection classification

Each dataset consists of data from 60 participants (30 Electromagnetic Hypersensitivity (EHS) and 30 Non-EHS) who participated in the 5G RF-EMF effect study and completed all four 5G BS signal exposures during pre-exposure, exposure, and post-exposure as outlined in Figure 2. The first dataset consists of Body Temperature, Blood Pressure and Pulse that were recorded before (Pre-Exposure), during (Exposure) and after (Post-Exposure) the assessment of 5G exposure. The Body Temperature, blood pressure and the pulse were recorded in Celsius (°C), millimeters of mercury (mmHg), and beats per minute (BPM) respectively. The second dataset was measured during exposure of 5G only. It consists of cognitive function component data parameters, which, were computed from the Psychology Experiment Building Language (PEBL) tests of Backward Digit Span Task (DSPAN) and Flanker Task, with outcome of Controlled for Accuracy (RT-A). Next, Berg's Card Sorting Task has three measured outcomes of Correct Percentage (C%), Percentage of Perseverative error (PE) and Percentage of Non-perseverative error (NPE). Lastly, the cognitive task named Tower of London Task has two outcomes, which are the Percentage of Success (S %) and the time needed until first move for each problem (FM). The physiological dataset involves 12 columns of normalized data parameters but for the analysis, the data is divided into each parameter based on the physiological parameter, which are Body Temperature recorded before 5G exposure (PreBT), the Body Temperature recorded during 5G exposure (ExpBT), the Body Temperature recorded after 5G exposure (PostBT), the Diastolic Blood Pressure recorded before 5G exposure (PreDIA), the Diastolic Blood Pressure recorded during 5G exposure (ExpDIA), the Diastolic Blood Pressure recorded after 5G exposure (PostDIA), the Systolic Blood Pressure recorded before 5G exposure (PreSYS), the Systolic Blood Pressure during 5G exposure (ExpSYS), the Systolic Blood Pressure recorded after 5G exposure (PostSYS), the Pulse recorded before 5G exposure (PreP), the Pulse/Heart rate recorded during 5G exposure (ExpP) and the Pulse/Heart rate recorded after 5G exposure (PostP). The raw data was first initially processed without any feature selection. This means that all the available features or attributes in the dataset were used as inputs for the classifier without any prior filtering or dimensionality reduction. The raw data was fed directly into the classifier, and the classifier made predictions or classifications based on this unaltered dataset. Next, the classifier predictions result was evaluated as shown in Table 2 and Table 3.

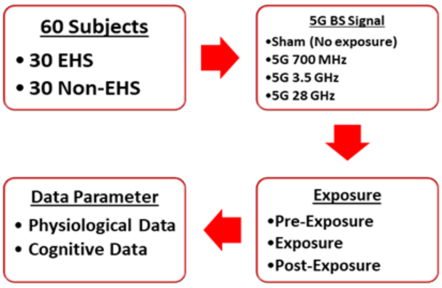

Fig.2. Dataset parameters used in this study.

Conventional single stage feature selection has the drawback of possibly selecting data after eliminating useful data during feature extraction stage. Thus, the raw data samples go through these stages to identify the best data normalization methods, best feature extraction methods and optimum hybrid features. For MSFS, the first stage consists of data reduction and data normalization methods. Data reduction through the removal of outliers after calculating their Interquartile Range (IQR) process involves identifying and eliminating data points that deviate significantly from the expected range of values within the dataset. The second stage consists of feature extraction methods, while the third stage and the fourth stage consist of feature selection and feature fusion, respectively. The outliers of any inconsistent data are cleaned and pre-processed to prepare the dataset for further processing.

Normalization methods are usually used to pre-process the data, to reduce the impact of differences in scale and units across variables, and to ensure that variables are comparable in a statistical analysis. In this study, 12 different normalization methods have been selected as shown in Table 1, including Z-score normalization (z score), Linear scaling (LS), Binary normalization (BNN), Bipolar normalization (BPN), Min-Max scaling (MMS), t-score normalization (t score), Decimal Inverse Logarithmic Scaled Normalization (DILSN), Relative Mean Normalization (RMN), Relative Standard Deviation Normalization (RSDN), Variation Normalization (VN), Robust Normalization (RN), and Relative Interquartile Normalization (RIN). These normalization methods have been evaluated based on their F-value and *p*-value analyses, leading to the identification of the top five normalization techniques in this research. These normalization formulas are defined as in the Equation 1 to Equation 12 and are calculated with the aid of MATLAB R2019 and Excel. The subsequent step involves performing a statistical t-test to identify the most notable *p*-value and the maximum F-value within each dataset. After the data has undergo first phase of MSFS and subjected to five distinct data normalization techniques. Then, followed by Principal Component Analysis (PCA) for the feature extraction stage in a subset of the dataset to calculate on the standardized data.

This step is to retain as much essential information as possible while reducing the dimensionality of the data. The outcomes include principal component coefficients, the transformed data in the principal component space, the eigenvalues of the covariance matrix, Hotelling's T-squared statistic content component. Next, for the feature selection in which involve choosing a subset of the most relevant features from the previous feature set. Combination of both techniques are employed to strike a balance between dimensionality reduction and information retention as well as avoiding overfitting. The integration of feature fusion technique is merging multiple sets of features into hybrid feature dataset which are their exposure data, type of subject and the cognitive and physiological dataset. This hybrid dataset is designed to encapsulate and represent the most valuable information from each of the contributing feature sets. The integrated hybrid feature dataset is the output result from the MSFS method in this study. The MSFS method plays a critical role in identifying and integrating the most relevant features from different feature sets, ultimately contributing to the creation of the hybrid feature dataset, which is a key component for achieving the study's objectives.

Table 1. Normalization method equation in pre-processing phase

No.	Equation	Type of Normalization Equation
1	$x' = \frac{x-\mu}{\sigma}$	ZS: z-Score Normalization Method equation, where x is the data input, μ is the mean data and σ is the standard deviation of the data.
2	$x' = \frac{(x-min)}{max-min}$	LS: Linear Scaling Normalization Method equation, where x is the data input, min is the minimum data and max is the maximum data.
3	$x' = \frac{0.8\,(x-min)}{max-min} + 0.1$	BNN: Binary Normalization Method equation, where x is the data input, min is the minimum data and max is the maximum data.
4	$x' = \frac{1.8\,(x-min)}{max-min} - 0.9$	BPN: Bipolar Normalization Method equation, where x is the data input, min is the minimum data and max is the maximum data.
5	$x' = \frac{x}{\mu}$	RMN: Relative Mean normalization, where x is the data input and μ is the mean data.
6	$x' = \frac{x}{\sigma}$	RSDN: Relative Standard Deviation Normalization, where x is the data input and σ is the standard deviation of the data.
7	$x' = \frac{x}{IQR}$	RIN: Relative Interquartile Normalization Equation, where x is the data input, IQR is the interquartile data.
8	$x' = \frac{x}{max-min}$	MMS: Min-Max Scaling Normalization Method equation, where x is the data input, min is the minimum data and max is the maximum data.
9	$x' = \frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$	TS: t - Score Normalization Method equation, where x is the data input, μ is the mean data, n is the number of total sample data and σ is the standard deviation of the data.
10	$x' = 10^{-12}10^{0.1x} * 10^{7}$	DILSN: Decimal Inverse Logarithmic Scaled Normalization Method equation, where x is the data input.
11	$x' = (x - median)/IQR$	RN: Robust Normalization equation, where x is the data input, and IQR is the interquartile range.
12	$C_{x,i} = \frac{\sigma}{\mu}\, x_i$	VN: Variation Normalization equation, where x is the data input, σ is the standard deviation of the data and μ is the mean data.

For the classification task, relevant features and ensuring on their input variables and output are selected. The Probabilistic Neural Network (PNN) classifier is trained and tested using a 90%/10% split of the data, where 90% of the data is used for training, and 10% is used for testing. K-fold cross validation of k=10 is chosen in the randomized dataset to obtain a more robust evaluation. For each iteration, the training data is prepared by selecting one group for testing and using the remaining groups for training. The training data is divided into input features and target outputs, and one-hot encoding is applied to the target outputs and the PNN trained using the training data. One-hot encoding is applied to the target outputs for classification tasks when the target labels data feature are categorical or nominal and can be used as features in ML models [30]. The PNN uses a non-parametric approach to estimate the probability density function of each class and then classifies new data points based on which class has the highest probability density at that point. The spread or the smoothing factor, σ in PNN is the calibration variable used to minimize the generalization error of the model. Varying σ gives control over the degree of nonlinearity of the decision boundaries for the network. The spread factor was set at 0.012 for cognitive dataset and 0.0135 for physiological dataset. A decision boundary approaches a hyperplane for large values of σ and approximates the highly nonlinear decision surface of the nearest neighbour classifier for values of σ that are close to zero.

Training Algorithm: The various steps involved in training PNN algorithm are described below,
Step 1: Data preparation for the input and output data class.
Step 2: Randomize the data to ensure unbiased training and testing.
Step 3: The randomized data is stored in a variable.
Step 4: Set the value of k=10 for number of folds in the cross validation, specify the number of rows in each fold for the grouping training data.
Step 5: Iterate over each group of data training and testing. The current group set as testing and remaining groups as training data.
Step 6: Extract the input features and output classes from training data.
Step 7: The classes are converted into binary vectors with ind2vec function.
Step 8: Spread factor is set as parameter to specific dataset for the classifier.
Step 9: PNN network is created using newpnn function and the input features and passed to the function.
Step 10: Extract the remaining data from the testing group as test inputs and test targets. Use the trained PNN to classify the test inputs and obtain the predicted outputs.

Result and Discussion

In order to compute the *p*-value and F-value for each parameter and ensure that the 5 best normalization methods will be selected for the design of stage two of hybridized MSFS by feature extraction method, feature selection, and feature fusion, which are very reliable for ML scope, the analysis is conducted to determine which normalization method should be used after the pre-processing stage in ML. PNN algorithm with from the outcome of the MSFS approach to classification that is well-suited to problems with a small to medium number of features. Evaluating the performance of PNN classifier involved using evaluation metrics to assess how well the classifier is performing in terms of classifying data points.

The primary metrics which result from the confusion matrix are used to evaluate the classifier's performance include accuracy, precision, F1-score, sensitivity, and specificity. The equations involved to obtain the evaluation metrics is the classified outputs are compared with the actual test targets and from the confusion matrix computed [31]. Equation (1) calculates the accuracy of the model, representing the correctness of the model's classification of data into their respective classes. Equation (2) measures precision, also known as positive predictive value, which indicates how well the model identifies positive cases accurately. Equation (3) quantifies recall, which measures the proportion of correctly predicted positive instances (true positives) out of all actual positive instances. It evaluates the model's ability to capture positive cases effectively. Equation (4) computes the F1-score, which provides a balanced measure of the model's performance by considering both precision and recall. It combines these metrics to assess overall performance. Equation (5) represents specificity, also known as true negative rate, which measures how well the model accurately identifies negative cases. Lastly, equation (6) corresponds to sensitivity, which is synonymous with recall. It measures the proportion of correctly predicted positive instances (true positives) out of all actual positive instances. Collectively, these equations offer a comprehensive set of metrics to evaluate and analyse the performance of a binary classification model across different aspects, including accuracy, precision, recall, F1-score, specificity, and sensitivity.

Table 2 and Table 3 tabulated the classification results to differentiate between subject involved of EHS and Non-EHS in the datasets as well as the exposure classification which consist of Sham, 5G 700 MHz, 5G 3.5 GHz and 28 GHz with and without the presence of MSFS. The model performance is severely affected from the proposed features in this case is the exposure data, subject data, and cognitive and physiological dataset. The approach of PCA is captured and capable when performing the feature selection techniques before utilizing in prediction ML models. Based on the prediction models performance classification as shown in Table 2 and 3 for the evaluation on the final model. First, we analysed the accuracy of all classification algorithms for two categories, separately. As for the classification process when raw data is passed directly to the classifier for both class categories (subject and exposure), accuracy values are less than 0.5 in Table 2. Using MSFS in the classification process significantly improves accuracy, with values greater than 0.94 for all algorithms and combinations of hybrid features. The PNN classifier performs optimally when the data has undergone MSFS processing before applying ML in Table 2. Similar evaluation performance is observed in Table 3 for the physiological dataset, with accuracy values improving when MSFS is applied. For both datasets, when data is processed directly to ML without MSFS, accuracy readings are low (less than 0.48). With the MSFS approach, accuracy values spike significantly, reaching more than 0.95. It was observed that for classification exposure, the normalization methods MMS and LS exhibited significantly increased specificity as well as accuracy. The normalization method named BNN for the category of subject classification accounted for most enhanced results in terms of specificity, accuracy, sensitivity, and precision. Next, it was shown that exposure classification with the normalization methods of MMS and LS featured higher data metrics of sensitivity and precision. For BCST (S), BCST (PE), and TOL (FM) data parameters, ZS normalization boasted the utmost level of specificity. From the physiological dataset results demonstrated that LS and BNN normalization methods achieved specificity, precision, and accuracy that were remarkably elevated. In the case of subject classification, the LS normalization method consistently achieved the most improved precision and specificity values. Using supervised ML techniques [32], this study demonstrated more profound insights into the features of data from short-term 5G base station exposure on the cognitive performance and physiological parameters of adults. In summary, the results suggest that incorporating MSFS process before applying ML algorithms enhances the accuracy of classification, especially in the context of both cognitive and physiological datasets. PNN classifier appears to be particularly effective when MSFS is employed. The comparison between direct ML application and MSFS demonstrates a substantial improvement in accuracy when using the feature selection approach.

(1) $$\text{accuracy} = \frac{\text{True}_{\text{positive}} + \text{True}_{\text{negative}}}{\text{True}_{\text{positive}} + \text{False}_{\text{positive}} + \text{True}_{\text{negative}} + \text{False}_{\text{positive}}}$$

(2) $$\text{precision} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{positive}}}$$

(3) $$\text{recall} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{negative}}}$$

(4) $$\text{f1} - \text{score} = \frac{2 \text{ x precision x recall}}{\text{precision} + \text{recall}}$$

(5) $$\text{specificity} = \frac{\text{True}_{\text{negative}}}{\text{True}_{\text{negative}} + \text{False}_{\text{ positive}}}$$

(6) $$\text{sensitivity} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{ negative}}}$$

By implementing MSFS method and use of PNN classifier on physiological and cognitive datasets, valuable insights can be gained about the relationship between variables, which can be used to develop predictive models for different outcomes. We have formulated a predictive strategy to investigate the potential prediction of short-term exposure to 5G base stations on the cognitive performance and physiological parameters of adults. This marks the initial benefit of a supervised ML approach for characterizing scenarios involving weak RF-EMF exposure in adults. As emphasized in [27], there is a pressing need for additional research in this domain to understand the potential impact and influence of specific RF-EMF features on prediction outcomes.

The ML method utilized in this study demonstrates high accuracy, precision, recall, and F1-score, indicating that the model accurately classifies data with minimal misclassifications. These metrics are used to evaluate the performance of classification models, where the model predicts class labels for input data. Therefore, a model that combines PCA with PNN and achieves more than 90% accuracy, precision, F1-score, and recall perform effectively and accurately predicts the classes of data points. However, it is important to note that the model's performance may vary for different datasets, and it should be evaluated on a validation set or in a cross-validation setting to ensure suitability for the data. From this research with the highest accuracy of 0.952 for both cognitive data and physiological data which served as a key indicator of the model's capability to make precise predictions in performance across these two distinct types of data underscored the robustness and generalizability of the model. Moreover, this study establishes a standard for future ventures in data analysis and predictive modelling, promoting the development of models that are even more precise and robust. It adds to the body of knowledge regarding the potential benefits of ML in the field of bioelectromagnetics. The findings from this analysis could enhance our comprehension of the specific data variables that should be collected in future research to elucidate the factors contributing to both high and low levels of weak RF-EMF exposures. With an expanded pool of experimental data in the future, the sample size can be increased, leading to more accurate outcomes. Future research can focus on exploring the effects of long-term exposure to 5G radiation on cognitive performance and physiological parameters, with a mechanism of action that involves alterations in neural activity or changes in hormone levels. In future work, it is suggested to increase the E-field intensity of millimeter wave signal exposure while ensuring it remains within the ICNIRP exposure limit for the public [33].

Table 2. Classification of subject and exposure result for cognitive data parameter

No	Data Parameter	Normalization Method	Classification	Presence of MSFS	Accuracy	Precision	F1-score	Sensitivity	Specificity
1	DSPAN	ZS, LS, RIN, RMN, RSDN	Subject	Yes	0.905	0.889	0.889	0.889	0.917
			Exposure		0.941	0.800	0.933	0.800	0.833
			Subject	No	0.477	0.479	0.530	0.479	0.474
			Exposure		0.222	0.219	0.246	0.452	0.463
2	FLANKER (RT-A)	BNN, MMS, DILSN, RIN, RMN	Subject	Yes	0.905	0.917	0.917	0.917	0.889
			Exposure		0.952	0.889	0.941	0.889	0.857
			Subject	No	0.488	0.489	0.626	0.489	0.483
			Exposure		0.230	0.235	0.299	0.475	0.476
3	BCST (C%)	ZS, BNN, BPN, DILSN, RSDN	Subject	Yes	0.952	0.923	0.960	0.923	0.909
			Exposure		0.952	0.833	0.923	0.905	0.750
			Subject	No	0.485	0.484	0.556	0.484	0.488
			Exposure		0.222	0.194	0.097	0.456	0.594
4	BCST (PE)	BNN, MMS, TS, RMN, RSDN	Subject	Yes	0.905	0.818	0.900	0.818	0.923
			Exposure		0.905	0.875	0.875	0.900	0.750
			Subject	No	0.498	0.498	0.558	0.498	0.499
			Exposure		0.228	0.230	0.337	0.465	0.456
5	BCST (NPE)	ZS, BPN, DILSN, RMN, RSDN	Subject	Yes	0.905	0.846	0.917	0.846	0.818
			Exposure		0.905	0.667	0.800	0.750	0.875
			Subject	No	0.478	0.473	0.275	0.473	0.480
			Exposure		0.227	0.227	0.236	0.476	0.463
6	TOL (S%)	BNN, MMS, DILSN, RMN, RSDN	Subject	Yes	0.952	0.923	0.960	0.923	0.909
			Exposure		0.952	0.833	0.909	0.800	0.952
			Subject	No	0.492	0.491	0.640	0.491	0.500
			Exposure		0.226	0.264	0.068	0.487	0.508
7	TOL (FM)	ZS, LS, MMS, RMN, RSDN	Subject	Yes	0.952	0.917	0.957	0.917	0.923
			Exposure		0.952	0.857	0.923	0.778	0.857
			Subject	No	0.483	0.491	0.594	0.491	0.457
			Exposure		0.228	0.236	0.315	0.490	0.409

Table 3. Classification of subject and exposure result for physiological data parameter

No	Data Parameter	Normalization Method	Classification	Presence of MSFS	Accuracy	Precision	F1-score	Sensitivity	Specificity
1	PreBT	ZS, BNN, DILSN, RMN, RSDN	Subject	Yes	0.952	0.923	0.960	0.923	0.923
			Exposure		0.952	0.800	0.889	0.857	0.857
			Subject	No	0.482	0.484	0.632	0.484	0.452
			Exposure		0.226	0.197	0.053	0.455	0.445
2	ExpBT	LS, BPN, MMS, TS, RSDN	Subject	Yes	0.952	0.923	0.960	0.923	0.889
			Exposure		0.952	0.889	0.941	0.889	0.800
			Subject	No	0.485	0.487	0.623	0.487	0.471
			Exposure		0.233	0.233	0.051	0.467	0.470
3	PostBT	ZS, LS, BNN, MMS, DILSN	Subject	Yes	0.952	0.900	0.947	0.900	0.923
			Exposure		0.905	0.875	0.875	0.875	0.778
			Subject	No	0.489	0.485	0.223	0.485	0.489
			Exposure		0.224	0.233	0.107	0.479	0.538
4	PreDIA	ZS, LS, BNN, MMS, RSDN	Subject	Yes	0.905	0.846	0.917	0.846	0.923
			Exposure		0.905	0.667	0.800	0.800	0.833
			Subject	No	0.475	0.479	0.485	0.479	0.471
			Exposure		0.223	0.274	0.120	0.597	0.455
5	ExpDIA	LS, BNN, BPN, MMS, TS	Subject	Yes	0.952	0.900	0.947	0.900	0.917
			Exposure		0.905	0.952	0.857	0.833	0.889
			Subject	No	0.479	0.475	0.507	0.475	0.485
			Exposure		0.228	0.225	0.231	0.481	0.480
6	PostDIA	BNN, BPN, MMS, RMN, RSDN	Subject	Yes	0.952	0.875	0.889	0.875	0.846
			Exposure		0.952	0.857	0.923	0.833	0.750
			Subject	No	0.473	0.471	0.616	0.471	0.497
			Exposure		0.231	0.211	0.124	0.462	0.474
7	PreSYS	LS, BNN, BPN, MMS, RSDN	Subject	Yes	0.952	0.917	0.952	0.867	0.909
			Exposure		0.952	0.875	0.933	0.875	0.857
			Subject	No	0.490	0.490	0.489	0.490	0.490
			Exposure		0.231	0.243	0.361	0.500	0.543
8	ExpSYS	ZS, BNN, BPN, MMS, RSDN	Subject	Yes	0.952	0.900	0.947	0.900	0.889
			Exposure		0.905	0.909	0.952	0.909	0.905
			Subject	No	0.476	0.475	0.514	0.475	0.478
			Exposure		0.223	0.201	0.196	0.431	0.359
9	PostSYS	BNN, BPN, MMS, TS, RMN	Subject	Yes	0.952	0.857	0.952	0.889	0.923
			Exposure		0.905	0.714	0.833	0.800	0.800
			Subject	No	0.488	0.481	0.533	0.481	0.497
			Exposure		0.228	0.232	0.357	0.478	0.500
10	PreP	LS, BPN, MMS, RMN, RSDN	Subject	Yes	0.952	0.929	0.941	0.905	0.923
			Exposure		0.952	0.857	0.923	0.857	0.800
			Subject	No	0.491	0.491	0.510	0.491	0.490
			Exposure		0.231	0.238	0.261	0.483	0.513
11	ExpP	LS, BNN, MMS, TS, RSDN	Subject	Yes	0.905	0.917	0.917	0.917	0.889
			Exposure		0.905	0.800	0.947	0.800	0.889
			Subject	No	0.481	0.484	0.569	0.484	0.475
			Exposure		0.221	0.224	0.318	0.464	0.452
12	PostP	LS, BNN, MMS, TS, RSDN	Subject	Yes	0.952	0.889	0.960	0.917	0.889
			Exposure		0.905	0.818	0.900	0.900	0.800
			Subject	No	0.479	0.480	0.486	0.480	0.478
			Exposure		0.236	0.222	0.107	0.459	0.434

Acknowledgement: *This research is funded under University-Private Matching Fund (UniPRIMA), Ref. No. UniMAP/RMC/9001-00702 (UniMAP Fund)/ 9030-00006 (FILPAL (M) Sdn Bhd Fund), and Tasneem Sofri under scholarship from JASSO Student Exchange Support Program with University of Yamanashi, Japan. The Malaysian Communications and Multimedia Commission (MCMC) approved the use data generated from the UniMAP research grant funded by MCMC (grant number: 2020/01/001) for this publication.*

Authors*: Tasneem Sofri, Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Arau 02600, Malaysia, E-mail: tasneem@studentmail.unimap.edu.my; Ts. Dr Allan Melvin Andrew, Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Arau 02600, Malaysia, E-mail: allanmelvin@unimap.edu.my; Associate Professor Ir. Ts. Dr. Hasliza A Rahim, Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Arau 02600, Malaysia, E-mail: haslizarahim@unimap.edu.my; Professor Dr. Hiromitsu Nishizaki, Integrated Graduate School of Medicine, Engineering, and Agricultural Sciences, University of Yamanashi, Kofu, 400-8510, Yamanashi, Japan, E-mail: hnishi@yamanashi.ac.jp; Associate Professor Dr. Latifah Munirah Kamarudin, Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, Arau 02600, Malaysia, E-mail: latifahmunirah@unimap.edu.my; Dr. Peng Wen Wong, Filpal (M) Sdn Bhd, Sains@USM, CREST Place Block A, 1st Floor, 10 Persiaran Bukit Jambul, 11900 Bayan Lepas, Pulau Pinang, Malaysia, E-mail: peng@filpal.com; Associate Professor Ts. Dr. Ping Jack Soh, Centre for Wireless Communications (CWC), University of Oulu, 90570 Oulu, Finland, E-mail: pingjack.soh@oulu.fi*

REFERENCES

[1] WHO, “Radiation: 5G mobile networks and health,” *World Health Orgnization (WHO)*, 2020. .

[2] C. L. Russell, “5 G wireless telecommunications expansion:

Public health and environmental implications," *Environ. Res.*, vol. 165, no. April, pp. 484–495, 2018, doi: 10.1016/j.envres.2018.01.016.

[3] WHO, "IARC Classifies Radiofrequency Electromagnetic Fields as Possibly Carcinogenic to Humans," *Press Release; World Heal. Organ.*, 2011, [Online]. Available: https://www.iarc.who.int/wp-content/uploads/2018/07/pr208_E.pdf.

[4] L. Redazione, "International Appeal: Scientists call for protection from non-ionizing electromagnetic field exposure," *Eur. J. Oncol.*, vol. 20, no. 3/4 SE-Specific topics, pp. 180–182, Dec. 2015, [Online]. Available: https://www.mattioli1885journals.com/index.php/EJOEH/article/view/4971.

[5] R. Nyberg and L. Hardell, "EU 5G Appeal – Scientists warn of potential serious health effects of 5G," *Jrs Eco Wirel.*, pp. 1–11, 2017.

[6] R. N. Kostoff, P. Heroux, M. Aschner, and A. Tsatsakis, "Adverse health effects of 5G mobile networking technology under real-life conditions," *Toxicol. Lett.*, vol. 323, no. January, pp. 35–40, 2020, doi: 10.1016/j.toxlet.2020.01.020.

[7] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran, and Q. Haq, "Machine Learning Techniques for 5G and Beyond," *IEEE Access*, vol. PP, p. 1, 2021, doi: 10.1109/ACCESS.2021.3051557.

[8] A. Haidine, F. Z. Salmam, A. Aqqal, and A. Dahbi, "Artificial Intelligence and Machine Learning in 5G and beyond: A Survey and Perspectives," in *Moving Broadband Mobile Communications Forward*, A. Haidine, Ed. Rijeka: IntechOpen, 2021.

[9] S. Khan *et al.*, "5G Vehicular Network Resource Management for Improving Radio Access Through Machine Learning," *IEEE Access*, vol. 7, pp. 1–9, Jan. 2020, doi: 10.1109/ACCESS.2020.2964697.

[10] S. Stanczak, S. Limmer, and L. Miretti, "Machine Learning for 5G and Beyond," *IEEE Access*, 2021, [Online]. Available: www.scopus.com.

[11] V. P. Rekkas, S. Sotiroudis, P. Sarigiannidis, S. Wan, G. K. Karagiannidis, and S. K. Goudos, "Machine Learning in Beyond 5G/6G Networks—State-of-the-Art and Future Trends," *Electronics*, vol. 10, no. 22, 2021, doi: 10.3390/electronics10222786.

[12] S. J. Regel *et al.*, "UMTS base station-like exposure, well-being, and cognitive performance.," *Environ. Health Perspect.*, vol. 114, no. 8, pp. 1270–1275, Aug. 2006, doi: 10.1289/ehp.8934.

[13] S. Andrianome, J. Gobert, L. Hugueville, E. Stéphan-Blanchard, F. Telliez, and B. Selmaoui, "An assessment of the autonomic nervous system in the electrohypersensitive population: a heart rate variability and skin conductance study.," *J. Appl. Physiol.*, vol. 123, no. 5, pp. 1055–1062, Nov. 2017, doi: 10.1152/japplphysiol.00229.2017.

[14] G. Curcio, E. Valentini, F. Moroni, M. Ferrara, L. De Gennaro, and M. Bertini, "Psychomotor performance is not influenced by brief repeated exposures to mobile phones.," *Bioelectromagnetics*, vol. 29, no. 3, pp. 237–241, Apr. 2008, doi: 10.1002/bem.20393.

[15] M. Koivisto, C. M. Krause, A. Revonsuo, M. Laine, and H. Hämäläinen, "The effects of electromagnetic field emitted by GSM phones on working memory.," *Neuroreport*, vol. 11, no. 8, pp. 1641–1643, Jun. 2000, doi: 10.1097/00001756-200006050-00009.

[16] C. Cinel, A. Boldini, E. Fox, and R. Russo, "Does the Use of Mobile Phones Affect Human Short-Term Memory or Attention?," *Appl. Cogn. Psychol.*, vol. 22, pp. 1113–1125, Dec. 2008, doi: 10.1002/acp.1425.

[17] A. Trunk *et al.*, "Effects of concurrent caffeine and mobile phone exposure on local target probability processing in the human brain," *Sci. Rep.*, vol. 5, p. 14434, Sep. 2015, doi: 10.1038/srep14434.

[18] H. Kleinlogel, T. Dierks, T. Koenig, H. Lehmann, A. Minder, and R. Berz, "Effects of weak mobile phone - Electromagnetic fields (GSM, UMTS) on well-being and resting EEG," *Bioelectromagnetics*, vol. 29, no. 6, pp. 479–487, 2008, doi: 10.1002/bem.20419.

[19] Z. Vecsei, B. Knakker, P. Juhász, G. Thuróczy, A. Trunk, and I. Hernádi, "Short-term radiofrequency exposure from new generation mobile phones reduces EEG alpha power with no effects on cognitive performance," *Sci. Rep.*, vol. 8, no. 1, p. 18010, 2018, doi: 10.1038/s41598-018-36353-9.

[20] S. Eltiti *et al.*, "Does short-term exposure to mobile phone base station signals increase symptoms in individuals who report sensitivity to electromagnetic fields? A double-blind randomized provocation study.," *Environ. Health Perspect.*, vol. 115, no. 11, pp. 1603–1608, Nov. 2007, doi: 10.1289/ehp.10286.

[21] J. Wallace, S. Andrianome, R. Ghosn, E. S. Blanchard, F. Telliez, and B. Selmaoui, "Heart rate variability in healthy young adults exposed to global system for mobile communication (GSM) 900-MHz radiofrequency signal from mobile phones," *Environ. Res.*, vol. 191, p. 110097, 2020, doi: https://doi.org/10.1016/j.envres.2020.110097.

[22] G. Oftedal, A. Straume, A. Johnsson, and L. J. Stovner, "Mobile phone headache: a double blind, sham-controlled provocation study.," *Cephalalgia*, vol. 27, no. 5, pp. 447–455, May 2007, doi: 10.1111/j.1468-2982.2007.01336.x.

[23] I. van Moorselaar *et al.*, "Effects of personalised exposure on self-rated electromagnetic hypersensitivity and sensibility – A double-blind randomised controlled trial," *Environ. Int.*, vol. 99, pp. 255–262, 2017, doi: 10.1016/j.envint.2016.11.031.

[24] G. Curcio, M. Ferrara, L. De Gennaro, R. Cristiani, G. D'Inzeo, and M. Bertini, "Time-course of electromagnetic field effects on human performance and tympanic temperature.," *Neuroreport*, vol. 15, no. 1, pp. 161–164, Jan. 2004, doi: 10.1097/00001756-200401190-00031.

[25] F. Malek, K. A. Rani, H. A. Rahim, and M. H. Omar, "Effect of Short-Term Mobile Phone Base Station Exposure on Cognitive Performance, Body Temperature, Heart Rate and Blood Pressure of Malaysians," *Sci. Rep.*, vol. 5, no. 1, p. 13206, 2015, doi: 10.1038/srep13206.

[26] S. Eltiti *et al.*, "Short-Term Exposure to Mobile Phone Base Station Signals Does Not Affect Cognitive Functioning or Physiological Measures in Individuals Who Report Sensitivity to Electromagnetic Fields and Controls," *Bioelectromagnetics*, vol. 30, pp. 556–563, Oct. 2009, doi: 10.1002/bem.20504.

[27] M. N. Halgamuge, "Supervised machine learning algorithms for bioelectromagnetics: Prediction models and feature selection techniques using data from weak radiofrequency radiation effect on human and animals cells," *Int. J. Environ. Res. Public Health*, vol. 17, no. 12, pp. 1–27, 2020, doi: 10.3390/ijerph17124595.

[28] A. A. A. Halim *et al.*, "Optimized Intelligent Classifier for Early Breast Cancer Detection Using Ultra-Wide Band Transceiver," *Diagnostics*, vol. 12, no. 11, 2022, doi: 10.3390/diagnostics12112870.

[29] A. Elkhouly, A. M. Andrew, H. A. Rahim, N. Abdulaziz, M. F. A. Malek, and S. Siddique, "Data-driven audiogram classifier using data normalization and multi-stage feature selection," *Sci. Rep.*, vol. 13, no. 1, pp. 1–14, 2023, doi: 10.1038/s41598-022-25411-y.

[30] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *J. Big Data*, vol. 7, no. 1, p. 28, 2020, doi: 10.1186/s40537-020-00305-w.

[31] G. A. Ismaeel, "Machine Learning to Diagnose Breast Cancer," *Prz. Elektrotechniczny*, vol. 99, no. 1, pp. 10–12, 2023, doi: 10.15199/48.2023.01.02.

[32] H. Huang, A. Y. Chang, and C. Ho, "Using Artificial Neural Networks to Establish a Customer-cancellation Prediction Model," *Prz. Elektrotechniczny*, vol. 89, no. 1b, pp. 178–180, 2013.

[33] ICNIRP, "Guidelines for Limiting Exposure to Electromagnetic Fields (100 kHz to 300 GHz).," *Health Phys.*, vol. 118, no. 5, pp. 483–524, May 2020, doi: 10.1097/HP.0000000000001210.