International Journal of Engineering Works
https://www.ijew.io/
SSN-p: 2521-2419
ISSN-e: 2409-2770
Vol. 13, Issue 01, PP. 01-06, January 26
https://doi.org/10.5281/zenodo.18183364

Neural Network based Portable Spectrometer for Fluid Classification

Hassan Nawazish Rasool[1], Muhammad Zeb[2]
[1]School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST), Islamabad ,
[2]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah campus
hassannawazishrasool1@gmail.com[1], mzebkhan97@gmail.com[2]
Received: 05 December 2025, Revised: 29 December 2025, Accepted: 06 January

***Abstract*—**The spectrometer, a powerful analytical instrument, plays a vital role in the fields of Chemistry, Material Sciences, Biochemistry, and Physics by allowing researchers to test, discover, and measure the spectral content of various fluids. This paper details the development of a portable, lightweight, cost-effective spectrometer device, which incorporates artificial intelligence for fluid classification. In a practical application, milk is used as the test fluid, with different qualities achieved by introducing controlled impurities. The spectrometer collected spectral data using a camera sensor, and the dataset is subsequently randomized and divided into training and testing data. Employing a range of machine learning algorithms and neural networks, the device accurately predicts the class of the fluid. The integration of optical components and microcontrollers facilitated model deployment. Notably, this device provides real-time fluid classification and displays results on an Organic LED. Beyond milk, its versatility allows for quality analysis of various fluids containing impurities, such as gasoline, human blood, and saliva. Remarkably compact at 70 x 70 x 50 mm and lightweight at 0.2 kilograms, the device had achieved an impressive average accuracy of 93.45 percent. It stands out for its ease of use, recharge ability, accuracy, and cost effectiveness compared to traditional spectrophotometers, positioning it as a valuable tool in the realm of scientific research and quality assessment.

Graphical Abstract

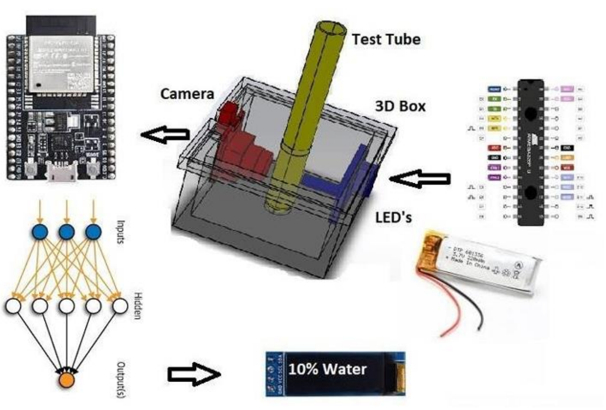

Figure 1. Core Components of Portable Spectrometer

***Keywords*—** AI based spectrometers, Fluids Classification, Smart Spectrometers.

I. INTRODUCTION

A spectrophotometer is an instrument used to measure the amount of light absorbed by a sample. It operates by directing a light beam through the sample to assess light absorbance[1]. Spectrometry finds widespread use in fields such as Chemistry, Biochemistry, Molecular Sciences, Biomedical Sciences, and Physics [2-4] . While wireless, cost-effective, and portable spectrometers are readily available for various measurements, they lack the capability to classify based on collected data [5] . Conversely, deploying an industrial grade spectrometer for local use is often infeasible. To address this, numerous devices have been developed for fluid quality assessment. Portable devices employing spectrometry, like NeoSpectra (Si-Ware, Cairo, Egypt) and SCiO (Consumer Physics, Herzliya, Israel), are employed for milk analysis. These devices utilize infrared radiation and mathematical modeling for predictions. For instance, SCiO offers a wavelength range from 740 to 1070 nm, while NeoSpectra's spectra range spans from 1350 to 2558 nm. Mathematical techniques used in these devices encompass Principal Component Analysis (PCA), cluster analysis, Partial Least Squares Regression, Savitzky-Golay derivatives, and the root mean square error of cross-validation for prediction [6,7]. These devices are recognized for their accuracy and portability. The application of spectroscopy for biodiesel quality assessment entails the examination of biodiesel's spectral attributes to infer its chemical composition and inherent characteristics. Spectroscopic techniques offer the advantage of swift and non-destructive quality evaluations for biodiesel[8] . In the realm of vehicle testing systems, spectrometers are employed to ascertain fuel quality control. This is pivotal in preventing engine damage and optimizing engine efficiency, especially in the presence of low-quality fuels such as naphtha, gasoline, diesel, and kerosene. Algorithms such as Support Vector Regression (SVR), Partial Least Squares (PLS), and Partial Least Squares Discriminant Analysis (PLS-DA) are utilized to classify these fluids effectively[9,10] . Fluid classification using Artificial Intelligence (AI) methods revolves around the training of machine learning models to recognize and categorize various fluids based on their inherent properties, characteristics, and patterns. AI techniques streamline and enhance the fluid classification process by efficiently analyzing extensive datasets and generating highly accurate predictions. The applications of AI-based fluid classification encompass the identification of diverse oil types, the detection of anomalies in hydraulic fluids, the characterization of chemical solutions, and more. Nonetheless, it's imperative to acknowledge that the success of AI-based fluid classification critically hinges upon the quality and diversity of the training data, as well as the judicious selection of appropriate algorithms and robust validation techniques [11]. The purpose of this study was to develop an affordable, user-friendly, portable photospectrometer for real-time fluid classification, employing Artificial Neural Networks (ANNs). The study also aimed to define the criteria for distinguishing between different fluid types and assessing their purity.

II. Hardware Description

The device is based on TSL1401 Linear sensor array-based camera module. The Parallax TSL1401-DB Camera Module in based on CMOS linear sensor array of 128 pixels and a lens of 7.9 mm. The module is easy to use and can be controlled by microcontrollers[12] . The camera is used for data acquisition using serial USB interface via PC. The device is able to measure the absorbance in the range within 300 nm to 1000 nm and it uses multiple LED's as light sources. A plot is shown in figure 2 describing the camera's range. The LED's used in the device are Red, Green, Blue, Infra-red and Ultraviolet to extract different features from the fluids. The spectrometer consists of 5 components, the LED's, TSL1401 Camera module, microcontrollers, power supplying battery and a displaying screen as shown in figure1. The 3D printed case in used to avoid the external light entering and the slot is available for insertion of glass tube (radius =15 mm) which contains the sample for measurement and classification. • The recorded data can be easily viewed and analyzed on the computer in CSV files. • The device is using multiple light source combination and use all data for the real time classification which is very useful and attractive feature in the spectrometer. • The size of the device is small as compare to other spectrometers and it is portable, rechargeable and easy to use anywhere for real time fluid classification. • The device is using the wide range of visible and nonvisible spectrum from 300 to 1000 nm and the visible and nonvisible light sources are used for data extraction which provides more details of the fluid features for classification.

A. Design Files

1) Mechanical Design

The 3D CAD design was crafted using SolidWorks, yielding an assembled device with overall dimensions of 75x50x50 mm. The assembled casing comprises two integral components: the primary body and the top enclosure, which

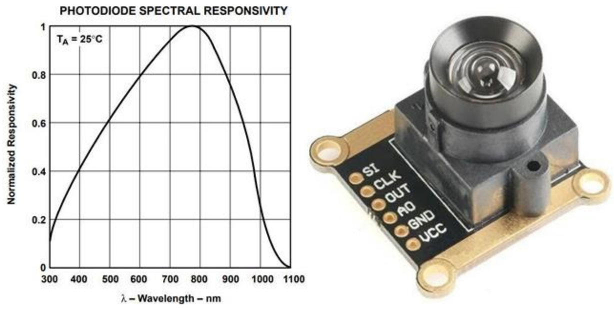

Figure 2. Spectral Range of Camera Sensor

accommodates the tube insertion area. Within the primary body, we house the camera module, a glass tube containing the test fluid, and LEDs that are affixed to a compact Printed Circuit Board (PCB). This assembly features apertures for securing screws employed in the attachment of the camera within the 3D-printed enclosure. A silicone adhesive is used to affix the circuit hosting the LEDs, while the camera is securely fastened with screws. The closure of the enclosure is a straightforward process facilitated by the top cover, effectively minimizing any unwanted light intrusion once sealed.

All the parts of the design are printed separately and there is no moving component in the device. Polylactic acid material is used for 3D printing. The material is light weight, easily available and having good mechanical strength. The ANET 3D printer is used to print the design and the 7 hours' time is taken by the printer for manufacturing. The CAD Design in isometric view is shown in Figure 3.

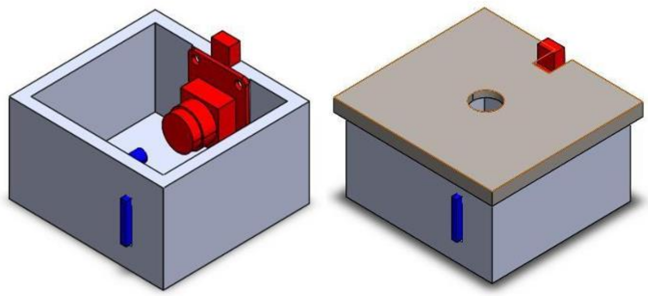

Figure 3. CAD Design created by Solid Works

2) Electrical Design

The schematic Diagram of the components is shown in figure 4. The device consists of multiple LED's (UV, IR and RGB), TSL1401 Camera Module, ESP-32, ATMEGA 328-P IC, Lithium polymer battery, buttons and OLED (128 x 32) for showing the labels. The electronic circuit schematic and the Printed Circuit Board (PCB) is designed by using the Eagle CAD software.

There is total two circuit boards one circuit board is inside the 3D box which have the LED's soldered on it. The internal circuit board design is shown in figure 5. The internal circuit board have LED's, resistors on top layer and a female header which is soldered on the bottom layer and inserted into the slot in box for providing the connections to the external PCB. The external PCB includes the slots and connections for ESP32 and ATMEGA 328P as shown in Figure 6.

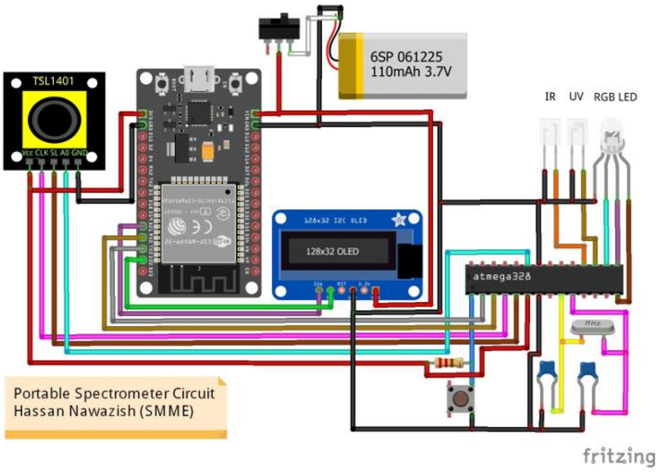

Figure 4. Circuit Diagram created by using Fritzing

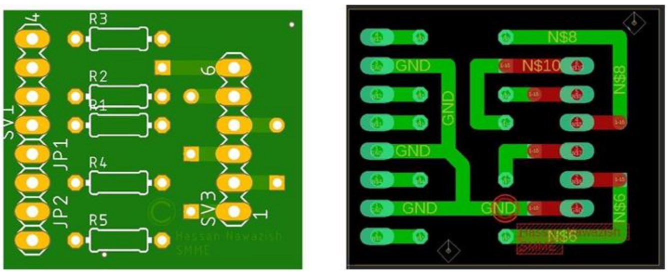

Figure 5. Eagle CAD schematic of Internal Printed Circuit Board

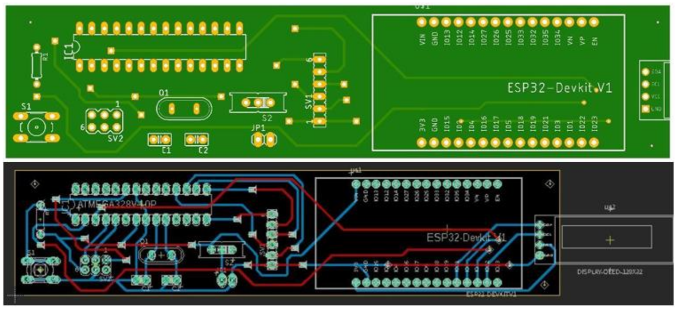

Figure 6. Eagle CAD schematic of External Printed Circuit Board

III. Methodology

A. Data Acquisition

The device can be used for fluid classification for testing the device milk has taken as fluid. The data is acquired by using the milk sample and the impurities such as water is added into the milk. The controlled amount of water is added into the milk using the scale. Various proportions of water were introduced into milk samples, and data acquisition was executed using MATLAB 2020. Subsequently, both impure and pure milk samples were introduced into a glass tube, which was then inserted into the testing apparatus. Data capture was conducted under conditions both with ambient lighting and in a controlled dark environment. The recorded data was meticulously stored in a CSV file using MATLAB. It is noteworthy that the raw signal exhibited noise stemming from light reflections and diffractions at the glass tube's edges. The noisy data was filtered, and the intensity of light passed through the glass tube's fluid was recorded by using the TSL1401 sensor array. There are total 128 small sensors in TSL1401 and the data from the sensor 28 to sensor 77 is extracted from the whole data as cleaned data. As there are total 5 light sources such as RED, Green, Blue, Ultraviolet and infrared so there are total 5 arrays.

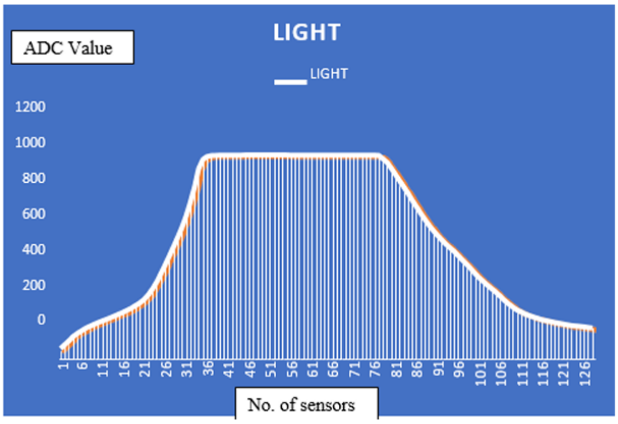

Figure 7. Raw Data

The raw data of the ADC when there is pure water in a glass tube is shown in the figure 7 as the light is reflecting and converging due to glass tube. Hence the data between the sensor 55 and sensor 77 is filtered out in the code for better analysis.

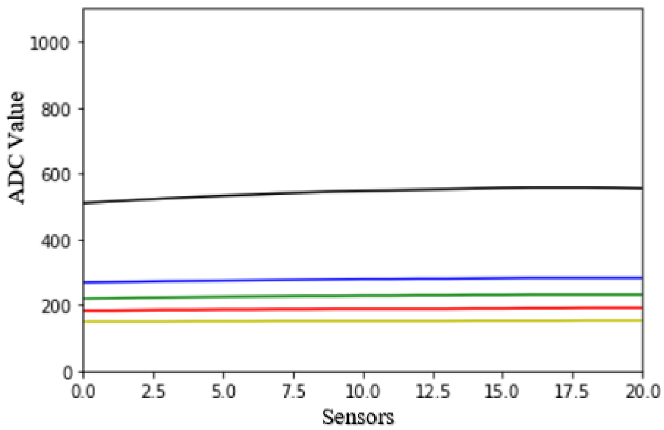

Figure 8. Filtered Data Samples

The sample of actual filtered data which is used to train the algorithm is given in figure 8. The data is taken and divided into 4 classes. Class zero is showing the Pure milk, class 1 is showing the milk with 20% water, class 2 is showing the milk with 40 % water and the class 3 is showing the milk with 50 % water. The data is raw Analog to Digital Converter (ADC) data. The 8-bit ADC is used and the range of the values are from 0 to 1024. When the light is passed through the fluid, the absorbance is different for different light sources as shown in figure 8. So, we have extracted multiple features from the fluid using multiple light sources.

B. Data Visualization

The overall data is visualized by using the Python Library called Seaborn[13] . The data is shown and labeled according to classes. Figure 9 is representing the box plots of the data.

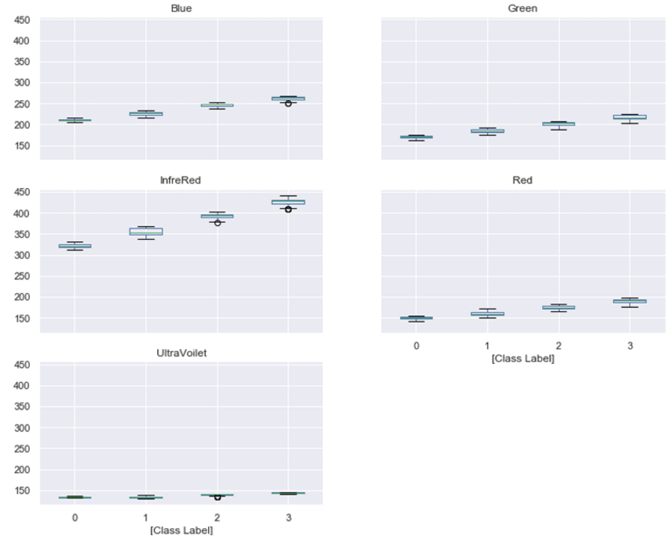

Figure 9. Data Visualization using Box Plots

C. Device Calibration and Validation

The TSL1401 camera sensor module exhibits a linear response characteristic. Notably, sensor calibration need not be performed for each distinct fluid under examination. Instead, the procedure involves data acquisition for each individual light source separately. The recorded output data was preserved in the form of raw Analog to Digital (ADC) values for each light source. Camera sensor calibration was executed by utilizing a conventional UV-Vis Spectrophotometer, which serves as a standard spectroscopy machine. The calibration process entailed the placement of the fluid within a cuvette, which was then positioned within the spectroscopy machine. Identical samples were utilized in both the standard machine and the portable spectrometer, and readings, encompassing both the Actual ADC values and the Expected ADC values, were aligned through adjustments in lens settings and potentiometer settings on the TSL1401 camera sensor module. The calculation of fluid absorbance was based on the application of the Beer-Lambert Law. Specifically, the following equation of the Beer-Lambert Law was employed to compute the absorbance.

$$A = \log_{10} \frac{I}{I`} \qquad (1)$$

Where, I` is the Incident Intensity.

I is the transmitted Intensity.

$$A = \varepsilon c I \qquad (2)$$

Where,

ε is the molar absorption coefficient.

C is molar concentration.

I is the optical length path.

D. Machine Learning

After data visualization, the data is divided into testing data and training data. The overall data is divided into 75% training and 25% testing data. After divided the data into the testing and training parts the data was randomized using the randomization script. The training data is used to train multiple Machine learning algorithms. Initially Support Vector Machine algo rithm was used for training and predicting the results. The results prediction after training the alg orithm without parameter tuning and with parameter tuning are 55.76 % and 92.307 % respective ly using support vector machine classifier. These accuracies are calculated by using the testing data. The classification results are shown in figure 10 for tuned support vector machine classifier. The classification results are shown for all five features.

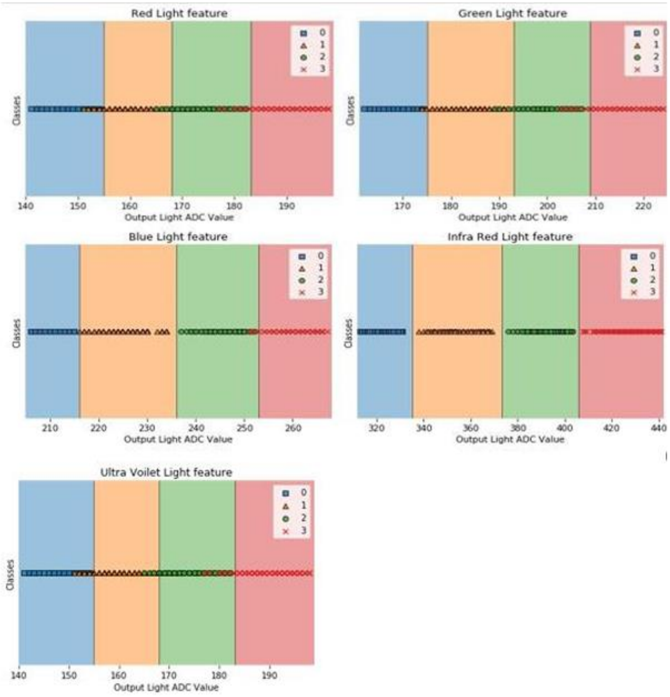

Figure 10. Support Vector Machine Classification Results

Other machine learning models are also used for best results such are Decision tree classifier, Gradient Boosting Classifier, Nearest Neighbor Classifier, Random Forest classifier and Logistic Regression. The train score and test score of all these classifiers are given in the Table 1.

The test scores, as reported, exhibit suboptimal performance, with a significant drawback associated with machine learning algorithms being the notable fluctuations in accuracy when subjected to continuous data randomization[14] . Notably, these fluctuations in accuracy persist with each iteration of data randomization.

In response to this challenge, the study leverages the deployment of an Artificial Neural Network (ANN). The ANN demonstrates commendable performance, yielding a consistent testing accuracy of 95%. Furthermore, the accuracy exhibits robustness, as indicated by a narrow tolerance range of ±0.8, even in the face of ongoing data randomization. The Figure 11 shown below is representing how accuracy increased with epochs. The overall workflow of the process the shown in Figure 12.

Table 1: Machine Learning Based Results

Classifier	Train Score	Test Score	Train Time(s)
Decision Tree	0.9968	0.6826	0.002
Gradient Boosting Classifier	0.9872	0.6057	0.458
Nearest Neighbors	0.8370	0.6057	0.001
Random Forest	0.9808	0.5673	0.084
Logistics Regression	0.8146	0.4711	0.009
Neural Network	0.9507	0.9428	

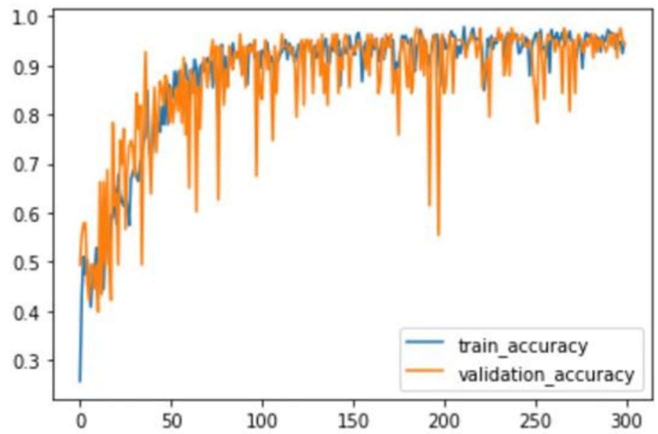

Figure 11. Training and Validation Accuracy of the Model

The implementation of machine learning algorithms is facilitated through the utilization of the scikit-learn Python library. Conversely, the training of the Deep Neural Network harnesses the capabilities of the Keras and TensorFlow Python libraries. The initial script, developed in Python, is subsequently transformed into a C++ script to be compatible with the Arduino Integrated Development Environment (IDE), thereby enhancing computational efficiency. Remarkably, the device presents real-time data visualization capabilities through an Organic LED (OLED) screen, underscoring its capacity for immediate results presentation.

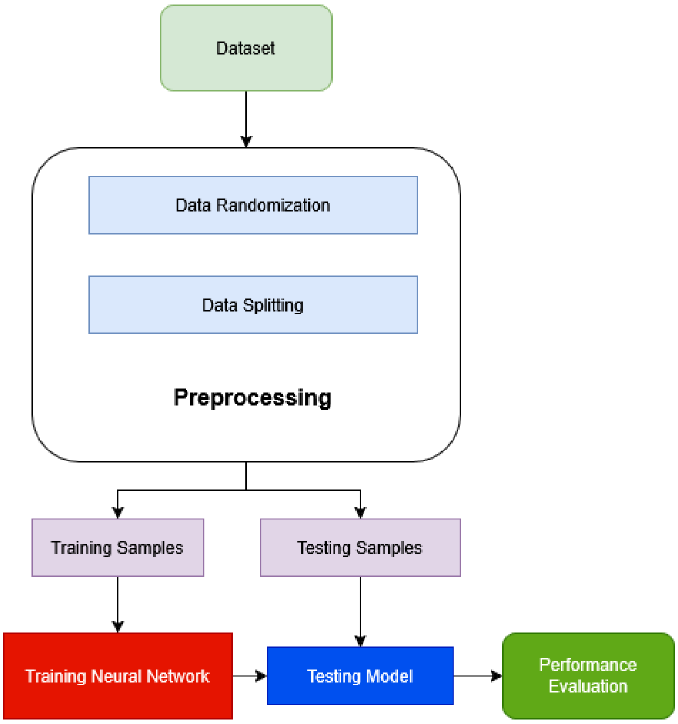

Figure 12. Process Workflow

Conclusion

The aim of this study was to develop a device which can be able to classify the fluid accurately and at real time. After this study we were able to classify the fluid (milk) with 95.5 % accuracy. And the device is extracting the features from the fluid using multiple light sources. Initially some machine learning techniques such as Support Vector Machine (SVM), K nearest neighbor (KNN) Decision tree classifier, Gradient Boosting Classifier, Nearest Neighbor Classifier, Random Forest classifier and Logistic Regression were used but the accuracy of prediction was changing whenever the data was randomized. Then Neural Network is used to classify the data which performs significantly better as compared to other machine learning techniques and provides the accuracy of 95.5 %, having ±0.8% tolerance after data randomization. The algorithm was trained and converted to C++ script to use in ESP32. OLED, Multiple light sources, TSL-1401 Camera sensor, ESP32 and ATMEGA 328P were used for training, predicting, and visualizing the results. Previous research has revealed the existence of devices employing similar techniques for fluid testing. Notably, certain devices utilize merely one or two light sources as discriminative features, achieving 100% accuracy in some cases, while others fall short of fluid classification capabilities. In contrast, the device in question employs a comprehensive set of five features, resulting in markedly enhanced accuracy. It is well-established that a greater number of critical features play a pivotal role in robust result prediction[5, 15] . Moreover, this device is distinguished by its cost-effectiveness and lightweight design when compared to commercially available milk classification devices like NeoSpectra (Si-Ware, Cairo, Egypt) and SCiO (Consumer Physics, Herzliya, Israel). This portable spectrophotometer demonstrates an admirable capability to classify fluids with precision and is economically priced, thereby offering substantial utility in monitoring the sale and quality control of synthetic milk, fuel, and various raw fluids. The dimensions of this device are compact, measuring 10 x 10 x 10 units, and its operation is simplified to the pressing of a single button, rendering it user-friendly. Additionally, the device is equipped with a rechargeable Lithium Polymer Battery, facilitating convenient charging. and light weight[16] . And it is useful as a simple test kit for Laboratory testing to train and test any fluid regularly. The successful incorporation of artificial intelligence for fluid classification introduces a new dimension to spectral analysis. Notably, the device's applicability extends beyond milk, making it a versatile tool for assessing the quality and composition of various fluids with impurities, such as fuel and biological samples. Its impressive accuracy, coupled with a compact and lightweight design, positions it as a cost-effective solution that holds great promise for scientific research and quality control applications. This innovation stands as a testament to the continued evolution of spectroscopy and its practical relevance in diverse scientific fields.

REFERENCES

[1] Mettler Toledo GmbH, "UV / VIS Spectrophotometry," Mettler-Toledo Int., no. September 2015, p. 56, 2021, [Online]. Available: https://www.mt.com/es/es/home/library/guides/laboratory-division/1/uvvis-spectrophotometry-guide-applications fundamentals.html

[2] F. Sánchez Rojas, C. Bosch Ojeda, and J. M. Cano Pavón, "Spectrophotometry | biochemical applications," Encycl. Anal. Sci., no. April, pp. 205–213, 2019, doi: 10.1016/B978-0-12-409547- 2.00501-1.

[3] P. Carpentier, A. Royant, J. Ohana, and D. Bourgeois, "Advances in spectroscopic methods for biological crystals. 2. Raman spectroscopy," J. Appl. Crystallogr., vol. 40, no. 6, pp. 1113–1122, 2007, doi: 10.1107/S0021889807044202.

[4] R. A. Ahmadi, F. Hasanvand, G. Bruno, H. A. Rudbari, and S. Amani, "Synthesis, spectroscopy, and magnetic characterization of copper(II) and cobalt(II) complexes with 2-amino-5-bromopyridine as ligand," Russ. J. Coord. Chem. Khimiya, vol. 39, no. 12, pp. 867–871, 2013, doi: 10.1134/S1070328413110018.

[5] K. Laganovska et al., "Portable low-cost open-source wireless spectrophotometer for fast and reliable measurements," HardwareX, vol. 7, p. e00108, 2020, doi: 10.1016/j.ohx.2020.e00108.

[6] J. Riu, G. Gorla, D. Chakif, R. Boqué, and B. Giussani, "Rapid analysis of milk using low-cost pocket-size NIR spectrometers and multivariate analysis," Foods, vol. 9, no. 8, 2020, doi: 10.3390/foods9081090.

[7] J. A. Diaz-Olivares, I. Adriaens, E. Stevens, W. Saeys, and B. Aernouts, "Online milk composition analysis with an on-farm near-infrared sensor," Comput. Electron. Agric., vol. 178, no. September, p. 105734, 2020, doi: 10.1016/j.compag.2020.105734.

[8] L. L. Monteiro, P. Zoio, B. B. Carvalho, L. P. Fonseca, and C. R. C. Calado, "Quality Monitoring of Biodiesel and Diesel/Biodiesel Blends: A Comparison between Benchtop FT-NIR versus a Portable Miniaturized NIR Spectroscopic Analysis," Processes, vol. 11, no. 4, 2023, doi: 10.3390/pr11041071.

[9] R. M. Correia et al., "Portable near infrared spectroscopy applied to fuel quality control," Talanta, vol. 176, no. August 2017, pp. 26–33, 2018, doi: 10.1016/j.talanta.2017.07.094.

[10] F. D. Santos et al., "Discrimination of oils and fuels using a portable NIR spectrometer," Fuel, vol. 283, no. August 2020, p. 118854, 2021, doi: 10.1016/j.fuel.2020.118854.

[11] I. U. Khan et al., "Amniotic Fluid Classification and Artificial Intelligence: Challenges and Opportunities," Sensors, vol. 22, no. 12, 2022, doi: 10.3390/s22124570.

[12] F. R. Fuentes, M. Trujillo, and C. Padilla, "Using Parallax TSL1401 DB Linescan Camera Module for line detection," pp. 1–17, 2011.

[13] A. Oberoi and R. Chauhan, "Visualizing data using Matplotlib and Seaborn libraries in Python for data science," Int. J. Sci. Res. Publ., vol. 9, no. 3, p. p8733, 2019, doi: 10.29322/ijsrp.9.03.2019.p8733.

[14] M. Kang, B. G. Ragan, and J. H. Park, "Issues in outcomes research: An overview of randomization techniques for clinical trials," J. Athl. Train., vol. 43, no. 2, pp. 215–221, 2008, doi: 10.4085/1062- 6050 43.2.215.

[15] R. C. Chen, C. Dewi, S. W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," J. Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00327-4.

[16] J. G. Cerón-Bretón, R. M. Cerón-Bretón, M. Rangel-Marrón, M. Muriel-García, Cordova- QuirozA.V., and A. Estrella-Cahuich, "International Journal of Energy and Environment," vol. 5, no. 3, pp. 328–336, 2011.

Hassan Nawazish Rasool received the MS degree from School of Mechanical and Manufacturing Engineering(SMME), NUST. He has done his bachelors in Electrical Engineering. He is currently working as Embedded Software Engineer at Maanz AI pvt Limited in domain of Automotive AI. His research interest includes biomechanical engineering, Health care Mangement, Software Design and Architecture, Product designing and manufacturing, Physics based simulation and optimization.

How to cite this article:

Hassan Nawazish Rasool, Muhammad Zeb "Neural Network based Portable Spectrometer for Fluid Classification" International Journal of Engineering Works, Vol. 13, Issue 01, PP. 01-06, January 2026. https://doi.org/10.5281/zenodo.18183364.