

Non-intrusive Testing of Liquid Culture Medium using Online NIR Spectroscopy and Machine Learning for Qualitative Analysis

Benjamin Samuel^{a,3}, Connor Reintjes^{a,1,2}, Paola González Pérez^{a,2}, Shiza Hassan^{a,3},
Dr. Amin Reza Rajabzadeh^{a,4}

^a*McMaster University, W. Booth School of Engineering Practice and Technology, Hamilton, ON., Canada*

Abstract

Insufficient quality assurance is a major expense associated with laboratories that can result in contamination, poor sample integrity, and lead to lost time due to repetitive sample testing. To ameliorate these issues, NIR spectroscopy has been combined with machine learning in this approach to qualitatively analyze the composition of liquid cultures in a non-intrusive and online manner. This will allow laboratories to save time by identifying contamination as soon as it happens and proceeding accordingly, as opposed to finding out after a full protocol has been performed. The method to achieve this involved creating a handheld casing to house the NIR which would take spectra of the sample and pass it to a machine learning model that would then identify whether the sample is in a normal or contaminated state. In phase 1 of the experiment, the NIR housing was built and initial testing was conducted for both contaminated and non-contaminated states, with the contaminant being *Lactobacillus rhamnosus*. The results achieved indicate that the NIR was able to differentiate between contaminated and non-contaminated samples. To analyze these readings a 1D-CNN model will be trained with the data collected thus far in the second phase of the experiment. Moving forward, further testing will be performed using quartz cuvettes to achieve more accurate and precise results as well as plating of bacteria to generate a standard curve to achieve quantitative results.

Keywords: Bioprocess Monitoring, Machine Learning, Near-Infrared Spectroscopy (NIR), 1D Convolutional Neural Network (1D-CNN), Qualitative Analysis

1. Introduction

Quality assurance issues are responsible for repeated testing, misdiagnosis, improper treatment and unpredictable costs [8]. The contamination of samples is the most common problem experienced in microbial laboratories. The real-time monitoring of samples can represent a significant

¹Conceptualization

²Validation & Software

³Investigation & Methodology

⁴Principle Investigator & Capstone Supervisor

cost-saving solution to improve quality assurance practices in laboratory and industrial settings [1]. To improve these issues, Near Infrared (NIR) spectroscopy has been combined with deep learning to provide real-time analysis of sample quality. This online analysis increases the testing and result acquisition speed while remaining non-invasive to reduce further contamination of the samples [12]. This technology also symbolizes a financial advantage by reducing the loss of biological samples using low computational requirements that can measure and classify the analyzed data as required. A deep learning model will be used to optimize this analysis as it offers advantages such as automatically integrating feature extraction, and incorporating hundreds of layers with many parameters compared to artificial neural networks [5]. The NIR spectral region has a range from 700 nm to 2500 nm. It is primarily characterized by absorption bands corresponding to the overtone and combination bands of C–H, N–H, O–H, and S–H stretching Figure 1. Hence, the spectra of most compounds will show unique absorption peaks in the NIR spectral region. These peaks can be used for the quantitative and qualitative analysis of the compounds in gases, solids or liquids.

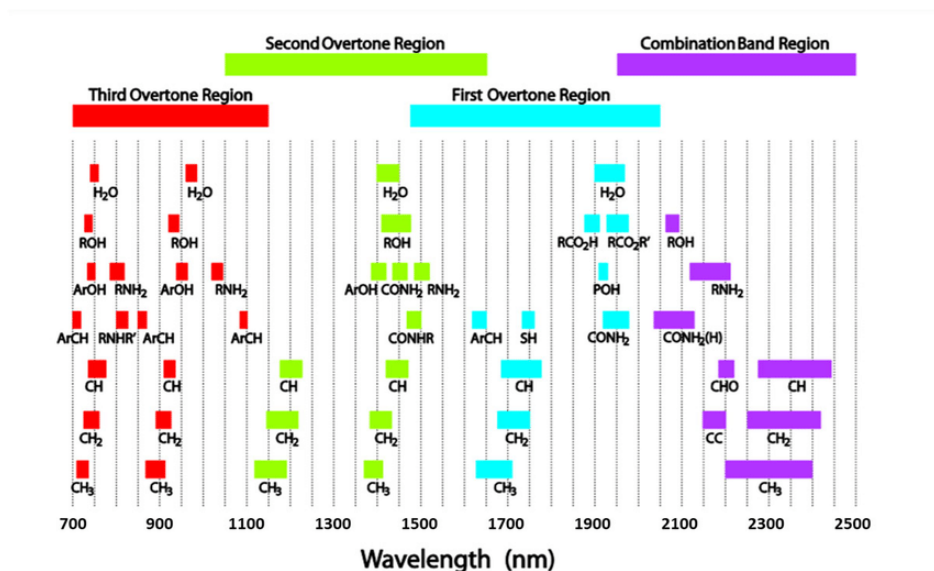


Figure 1: Principal analytical bands and relative peak locations for absorptions in the near-infrared region. Unique absorptions found in the majority of chemical and biological products can be utilized in both qualitative and quantitative examinations

Ethanol and water have characteristic NIR spectrum patterns that allow for their identification and quantification. Ethanol has a unique narrow peak at 1200 nm representing the second overtone of the C–H stretching vibration [2]. Water, on the other hand, shows a flat behaviour at this wavelength due to the lack of C–H bonds.

The intensity of NIR absorption bands is 10–100 times lower than that of the equivalent basic mid-IR absorption bands. This facilitates the direct analysis of strong absorbances and high light-scattering efficiency. Band overlap and penetration depth reduce in the near-infrared (NIR) spectral region, whereas the efficiency of light scattering and absorptivity improves with wavelength. As NIR spectroscopy depends on light absorption, spectral data can be acquired either in transmittance

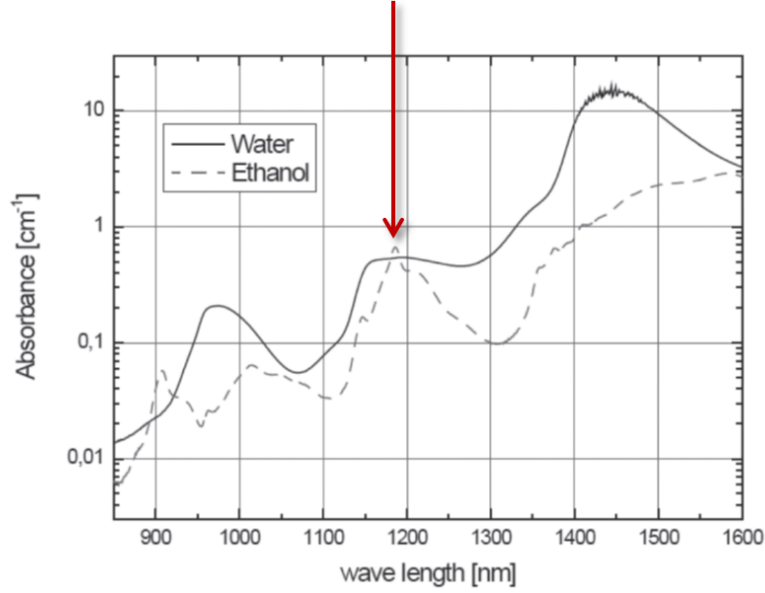


Figure 2: NIR spectrum of ethanol and water [2]

or reflectance mode. Diffuse reflectance ($R = \frac{I_r}{I_0}$) measurements are favoured for opaque or light-scattering matrices whereas translucent samples used transmittance ($T = \frac{I}{I_0}$). NIR spectroscopy's quickness, adaptability to a variety of materials, and capacity to examine liquid samples make it a promising tool for bio applications [3]. Recent advancements are making NIR spectroscopy more accessible with portable handheld instruments that allow for real-time, on-site examination.

Cell culture media (CCM) plays a vital role, directly impacting process yield and product quality [9]. CCM's primary objective is to create and sustain an optimal physiological environment for large-scale cell culturing, ensuring cell health and expression of desired Critical Quality Attributes (CQAs). However, it's crucial to recognize that the chemical and physical properties of CCM are sensitive to microbial growth, chemical reactions, and environmental factors like temperature and light. Without sterilization, CCM can rapidly change due to microbial contamination, leading to increased turbidity and light scatter, consequently elevating baselines and noise in spectra, potentially compromising measurements. While chemometrics can mitigate some measurement errors, identifying errors induced by samples and measurements in CCM analysis can be challenging.

NIR spectra are known to be high-dimensional data due to the baseline drift and other noise signals present. This results in the need for preprocessing techniques for noise filtering and dimensionality reduction before analyzing the true compound signals. Hence, NIR spectra usually involve the use of chemometric algorithms like partial least squares regression (PLSR) and support vector regression (SVR) to clean the data from baseline drift as well as to understand the chemical changes in the samples [10, 6]. These algorithms require trained individuals to obtain the necessary parameters used to analyze the spectra [14]. Similarly, chemometric models are often limited by their inability to generalize for variance in spectra from different instruments, or changing storage and growing conditions (P03). Deep Learning Machine Learning (ML) can be used to automate extracting the main compound features from high-dimensional spectral data and eliminate the need for feature-

selection techniques [13].

While different classes of ML algorithms have successfully achieved the classification of NIR spectra, traditional ML methods such as partial least squares (PLS), K-nearest neighbor (KNN), and principal component analysis (PCA) require a higher level of expertise to design suitable features of the model’s architecture [16]. Deep learning, on the other hand, does not require a high level of expertise as it utilizes the raw features of data to analyze and classify it as needed. This is achieved through multiple hidden layers that are trained end to end. Some of these layers can be specialized convolution layers that allow for the learning and identification of local feature patterns. This aspect enables deep learning architectures to employ less preprocessing for high-dimensional data [15].

Convolutional neural networks (CNNs) are a class of deep neural networks commonly used for data analysis due to their successful results in image processing and classification [11], speech recognition, and other computer vision tasks. CNNs are a feed-forward multi-layer architecture in which a kernel or filter takes specific features from local regions from the upper layer. This architecture allows for an autonomous extraction of important features from complex data for analysis and learning [16]. A common drawback of the use of CNNs for spectral analysis is the requirement of large data sets during the training process of the model. To avoid overfitting due to a limited number of data samples, one-dimensional CNN (1D-CNN) can be used. These models are similar to traditional CNNs except for the data size required and low computational requirements. 1D-CNNs have proven to have good information extraction and high classification accuracy with simple preprocessing techniques. It has been applied by [10] for the analysis and classification of NIR spectra from breast cancer tissue to aid in cancer diagnosis, demonstrating a 94.67% classification accuracy. Chai et al. [4] also developed an improved 1D-CNN structure to discriminate *Anoectochilus roxburghii* from its counterfeits using NIR spectra.

Previous studies performed on 1D-CNN for the analysis of NIR spectral data demonstrate the viability of its application for the online analysis of culture media. The present study aims to develop a method based on NIR spectra acquired on a portable device, as a non-destructive, online method to assess quality attributes of culture media. Due to the expected nature of the data collection, it was necessary to employ a time series analysis when designing the CNN structure. Within this experiment *S. cerevisiae* was used as the model, testing it under two conditions, normal and contaminated with Lactic Acid Bacteria. *Lactobacillus rhamnosus* was used as the contaminant of choice due to its improved growth in the presence of *S. cerevisiae*. Also, using this species ensures the proper growth of *S. cerevisiae* as it remains unaffected in its presence [7]. The NIR spectra of ethanol were analyzed in each case and passed to the deep learning model which classified the “normal” baseline against a “non-normal” sample.

2. Materials and Methods

2.1. Materials

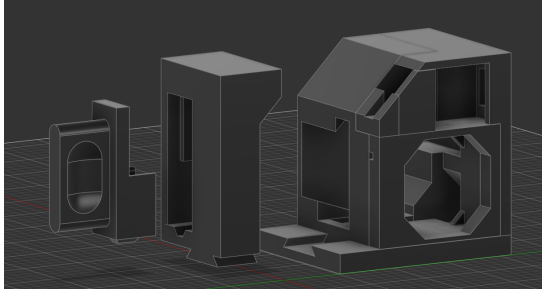
2.1.1. Material List

Item	Supplier	Catalog Number	Quantity	Price	URL
<i>S. Cerevisiae</i>	Sigma Aldrich	YSC1	100g	\$192 CAD	Link
Peptone	Sigma Aldrich	68727	5g	\$165 CAD	Link
Yeast Extract	Sigma Aldrich	Y6125	5g	\$119 CAD	Link
Dextrose	Sigma Aldrich	G7021	10g	\$74 CAD	Link
CytoSoft®T-25 Flasks	Sigma Aldrich	CC330	1x * 10	\$247 CAD	Link
ABS Filament	Bambu Lab	ABS Black w/ Spool	1x 1kg	\$35.99 CAD	Link
<i>Lactaseibacillus rhamnosus</i>	ATCC	7469	1 vial	—	Link

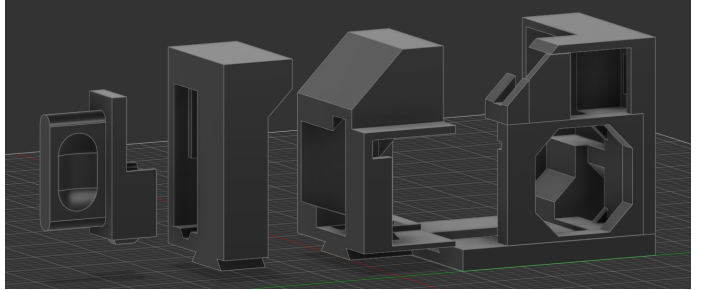
Table 1: Material Information Table with URLs

2.1.2. NIR Housing Design

Due to the design of the DLP NIRscan Nano (Texas Instruments) and the location of the sensor window on the unit, it required a housing to be designed. This housing was designed to both hold the NIR as well as the sample to ensure consistent scan conditions. The housing was created using a 3D printer (P1S, Bambu Labs) using Matte PLA filament (Bambu Labs) for the test models, and was reprinted in ABS filament (Bambu Labs) for the final housing. The housing underwent multiple iterations with all design and modeling performed within AutoDesk Fusion360 parametric CAD software. The final iteration of the housing used can be seen below in Figure 3. All of the components slide along a single dovetail mount along the bottom plate to allow for quick assembly while keeping all components secure. Due to slight problems with temperature regulation, a NF-A4x10 5V PWM fan from Noctua was added in addition to a fan speed controller (NA-FC1), to prevent the NIR from overheating during scans. A cuvette holder was also designed to allow for the quick change of samples, while also ensuring that the cuvette remained fixed in place for scans.



(a) NIR Housing



(b) NIR Housing Fully Exploded

Figure 3: NIR Housing Design

2.2. Methods

2.2.1. Ethanol Standard Curves

To confirm the proof of concept of the project, ethanol standard curves were created and detected using the NIR to determine the differences in signals between different concentrations of ethanol. 5 – 6 solutions of varying ethanol concentrations were tested in two stages, at higher concentrations and low concentrations. For the first standard curve, 300mL of distilled water was retained in a large beaker. Five, 100mLL beakers were set up and labelled according to the scheme in Table 2. These volumes were added to each beaker and 3mL were subsequently taken out and placed into a plastic cuvette. The cuvette was cleaned with a kimwipe and then placed in the NIR to take a reading. The process was repeated for all beakers 3 times to generate 3 replicates for each sample. The concentrations used were far enough so it would make it easier to tell a difference between the absorbance readings taken from the NIR.

Sample	Ethanol Volume (mL)	Distilled Water Volume (mL)
1	0	100
2	25	75
3	50	50
4	75	25
5	100	0

Table 2: Ethanol and Distilled Water Volumes for Initial Standards

For the second curve, smaller concentrations of ethanol were used to see if the NIR was able to detect small differences in concentrations of ethanol. The process was similar to the first standard curve. In this case, 600mL of distilled water was retrieved in a beaker and six, 100mL beakers were set up and labelled according to the scheme in Table 3. The remainder of the process was the same as before with the labelled volumes being added to each beaker and 3mL extracted to be placed in a plastic cuvette. The NIR spectra was taken and the process was repeated 3 times to generate 3 replicates.

Sample	Ethanol Volume (mL)	Distilled Water Volume (mL)
1	0	100
2	1	99
3	2	98
4	3	97
5	4	96
6	5	95

Table 3: Ethanol and Distilled Water Volumes for Low Concentration Standards

2.2.2. Yeast Culture

This section aims to prepare liquid cultures of yeast so that the classification model can be trained to detect contamination. 500mL of YPD was meticulously prepared to carry out this process successfully. 5g of Yeast extract, 5g of Peptone, 10g of Dextrose, and 500mL of dH₂O were mixed and autoclaved, ensuring the broth was sterilized. 30mL of autoclaved broth was aliquoted in a bio safety cabinet (BSC) and 0.06g of yeast was added to the broth. The liquid culture was re-suspended and ten 3mL samples were aliquoted into 15mL falcon tubes. The aliquots were then stored at -20°C, preserving the viability of yeast cells. The experimental procedure depended on whether the culture was inoculated with the contaminant.

For each run, the samples were thawed at 25°C for 10 minutes and then pipetted into the cuvette. 3mL of yeast culture was aliquoted into a cuvette and layered with 50μL mineral oil. This ensures that anaerobic fermentation takes place. For runs with contamination, 2.8mL of yeast culture, 200μL of *L. rhamnosus*, and 50μL of mineral oil were added into a cuvette. Varying concentrations of contamination were used and the procedure is expounded on below. The cuvette was covered with parafilm and placed in the NIR. The time intervals for data collection varied as technical issues persisted throughout the experiment.

2.2.3. Lactic Acid Bacteria Culture

A known contaminant has to be prepared and added to the yeast sample. Freeze-dried lactobacillus was utilized to create a bacterial liquid culture of 75mL. 74.5mL of YPD was added to an autoclaved flask in a BSC. 500μL of YPD was used to rehydrate 1 vial of *L. rhamnosus*. The rehydrated *L. rhamnosus* was added to the flask and then placed inside a shaking incubator at 37°C for 3 days.

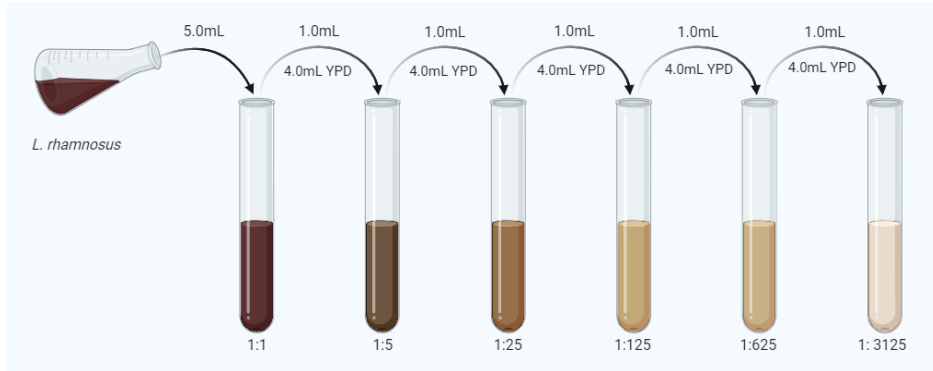


Figure 4: Lactic Acid Bacteria Serial Dilution

After 3 days, serial dilution was performed to create varying concentrations of bacteria to mimic the different possible levels of contamination. 5.0mL of bacterial culture was aseptically transferred to the first falcon tube. From the first tube, 1.0mL of bacterial culture was added into a second falcon tube and mixed with 4.0mL of YPD broth. The procedure was repeated to achieve concentration in the ratio of 1 : 5 until a ratio of 1 : 3125 was achieved.

2.2.4. Glycerol Freezing

The purpose of the glycerol freezing was to ensure the bacteria was preserved in a manner in which there would be limited growth and no death. This was done by letting the bacteria grow in the shaking incubator at 37°C for 7 days. Afterwards, 32mL were extracted from the sample and centrifuged at 10,000g at 4°C for 10 minutes in 2 50mL falcon tubes. 2 solutions of varying concentrations were created, with the first being a hyperconcentrated sample. This was done by extracting 17mL from the first tube resulting in a volume of 15mL + bacterial pellet. The pellet was resuspended in the 15mL by vortexing the sample. From this 3 solutions were created, 2 solutions of 5mL glycerol and 5mL sample, and 1 solution of 4mL glycerol and 4mL sample. For the second 32mL sample, none of the broth was removed to create a normal concentration. The pellet was resuspended in the 32mL by vortexing and 3 solutions of 5mL glycerol and 5mL sample. Afterwards, the tubes were all inverted and placed in a freezer at -20°C.

2.2.5. Spectral Measurement

A laptop running Windows was used in conjunction with the DLP NIRscan Nano GUI software from Texas Instruments to control the scans. Every sample was taken using a modified Hadamard profile Table 4. Before any spectra are taken, the software is first booted up and directories are updated for the new sample. A new folder is made for each sample with the naming convention "Sample #T". '#' is represented by the sample number in the sequence (1-5) and 'T' represents the sample type with 'C' being contamination and a blank value being the normal. Each spectra file is saved as a .csv which is named based on the sample name, a serial number, and a date/time stamp.

Setting	Value
Start Wavelength (nm)	900
End wavelength (nm)	1500
Pattern Pixel Width (nm)	4.68
Exposure (ms)	1.27
Digital Resolution	264
Scan Repeats	9
PGA Gain	64

Table 4: Ethanol and Distilled Water Volumes for Low Concentration Standards

The lamp of the NIR is first toggled on to warm it up and equalize the NIR. This is left on until the scanning sequence is ready to start. A blank sample of an empty cuvette is first taken to replace the factory background of the NIR. The sample is then loaded into the holder. The software is then set to take 196 scans every 900 seconds which equates to a single scan every 15 minutes for a total of 48 hours. The lamp is then turned off, and the first scan is then taken before leaving the apparatus.

3. Results

Results

3.1. Data Analysis

Analysis

4. Discussion

Discussion

5. Conclusion

Conclusion

Acknowledgments

Acknowledge any individuals or organizations that contributed to the research, funding bodies, and any supporting institutions.

References

- [1] E. Abatenh, B. Gizaw, and Z. Tsegaye. Contamination in a microbiological laboratory. *International Journal of Research Studies in Biosciences*, 2018.
- [2] Anton Paar GmbH. Near infrared spectroscopy (NIR) | anton paar wiki.
- [3] K. B. Beć, J. Grabska, and C. W. Huck. Near-Infrared Spectroscopy in Bio-Applications. *Molecules*, 2020.
- [4] Q. Chai, J. Zeng, D. Lin, X. Li, J. Huang, and W. Wang. Improved 1D convolutional neural network adapted to near-infrared spectroscopy for rapid discrimination of *Anoectochilus roxburghii* and its counterfeits. *Journal of Pharmaceutical and Biomedical Analysis*, 2021.
- [5] Mishra P et al. Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, 2022.
- [6] I. A. Naguib, E. A. Abdelaleem, M. E. Draz, and H. E. Zaazaa. Linear support vector regression and partial least squares chemometric models for determination of Hydrochlorothiazide and Benazepril hydrochloride in presence of related impurities: A comparative study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 130:350–356, 2014. doi: 10.1016/j.saa.2014.04.024.
- [7] Stefano Nenciarini, A. Reis-Costa, M. Pallecchi, et al. Investigating yeast–lactobacilli interactions through co-culture growth and metabolite analysis. *Fermentation*, 9(11):933–933, 2023. doi: 10.3390/fermentation9110933.
- [8] Radox Laboratories. The cost of poor quality in the clinical laboratory quality control.
- [9] A. G. Ryder. Cell culture media analysis using rapid spectroscopic methods. *Current Opinion in Chemical Engineering*, 2018.
- [10] H. Shang, L. Shang, J. Wu, Z. Xu, S. Zhou, Z. Wang, H. Wang, and J. Yin. NIR spectroscopy combined with 1D-convolutional neural network for breast cancerization analysis and diagnosis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2023.
- [11] N. Sharma, V. Jain, and A. Mishra. An analysis of convolutional neural networks for image classification. *Procedia Computer Science*, 132:377–384, 2018. doi: 10.1016/j.procs.2018.05.198.
- [12] S. I. Sohn, S. Pandian, Y. J. Oh, et al. An overview of near infrared spectroscopy and its applications in the detection of genetically modified organisms. *International journal of molecular sciences*, 22(18):9940, 2021. doi: 10.3390/ijms22189940.
- [13] J. Walsh, Arjun Neupane, Anand Koirala, et al. Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. II. The rise of convolutional neural networks. *Journal of Near Infrared Spectroscopy*, 31(3):109–125, 2023. doi: 10.1177/09670335231173140.
- [14] Y. Wang, Y. Jiang, and J. Lan. FCNN: An efficient intrusion detection method based on raw network traffic. *ResearchGate*, 2021. doi: 10.1155//2021//5533269.
- [15] W. Yue, J. Yiming, and L. Julong. A fast deep learning method for network intrusion detection without manual feature extraction. *ResearchGate*, 2021. doi: 10.1088//1742-6596//1738//1//012127.
- [16] Wenwen Zhang, Liyanaarachchi Chamara Kasun, Qi Jie Wang, Yuanjin Zheng, and Zhiping Lin. A Review of Machine Learning for Near-Infrared Spectroscopy. *Sensors*, 22(24):9764, January 2022. ISSN 1424-8220. doi: 10.3390/s22249764.