Saliva FTIR Spectra and Machine Learning for Autism Spectrum Disorder Diagnosis—Preliminary Study

Mayara Moniz Vieira Pinto, Emilia Angela Lo Schiavo Arisawa, Leandro José Raniero, and Tanmoy Bhattacharjee

***Abstract*—The diagnosis of Autism Spectrum Disorder (ASD) remains a challenge due to the lack of specific tests and biological markers. ASD is a neurodevelopmental disorder that affects individuals throughout their lives, and its diagnosis allows access to treatments that improve their prognosis. Saliva analysis by Fourier Transform Infrared Spectroscopy (FTIR), which was not previously reported, appears to be a promising diagnostic tool for ASD. This study acquired spectra from samples of 19 ASD and 19 control children. Spectral signatures suggest the dominance of protein secondary structures, β-pleated sheet and α-helix structures in ASD and control children, respectively. Support Vector Machine (SVM) gave the best diagnosis, with sensitivity, precision, and specificity being 92%, 94%, and 95%, respectively. Shapley values analysis to understand the impact of spectral features on the SVM classifier identified β-pleated and β-turn sheets as responsible for classification. Results indicate the potential of saliva-based FTIR for autism diagnosis, warranting a large-scale trial.**

***Index Terms*—Autism spectrum disorder, diagnosis, FTIR, machine learning.**

I. Introduction

AUTISM Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in communication, social interaction, and learning, with restricted and repetitive patterns of behavior, such as continuous movements, fixed interests, and hypo- or hyper-reactivity to sensory stimuli [1]. These characteristics vary in the form of manifestation and in the degree of intensity, influencing the way each person with ASD relates, expresses, and behaves.

In 2016, the Centers for Disease Control and Prevention (CDC) estimated that Caucasians and Afro-descendants had a prevalence of ASD of 1 in 54 children aged 8 years [2], with boys being affected four times more often than girls [3]. Early intervention for children with autism is key owing to common difficulties in communication.

A significant increase in the clinical diagnosis of ASD has been detected in recent decades, making it a major public health concern [4]. The diagnosis of ASD is difficult and complex due to the absence of specific tests and biological markers. The current diagnostic technique is the investigation and clinical observation of the individual's behavior undertaken by a multidisciplinary team. Late identification of ASD delays the initiation of specific therapeutic and pharmacological interventions, affecting the individual's biopsychosocial development and prognosis. Early diagnosis and intervention can reduce the need for educational and behavioral support in primary school and throughout life [5].

The suggested ASD etiology involves the interaction between genetic and environmental factors, as well as systemic and central nervous system (CNS) inflammation [6]. ASD individuals commonly present difficulty maintaining a balanced diet due to multiple factors, such as high selective food preferences, organic gastrointestinal (GI) diseases, and oral motor difficulties, presenting high rates of intestinal dysbiosis compared to neurotypical individuals. The authors also characterized the oral and intestinal microbiota of ASD patients comparing them to neurotypical individuals, identifying that the microbiotas differ between individuals with and without a diagnosis of ASD [7].

Studies indicate a significant correlation between the severity of GI dysfunction and the severity of symptoms, reporting that alterations in the intestinal microbiota can impact neurobehavioral function [8]. The intestinal microbiota (IM) is an important factor in the microenvironment of the digestive tract, and may influence symptoms, considering the observation of distinct intestinal microbiota in children with ASD compared to neurotypical ones [9].

Evidence from animal studies has supported the association between microbiota dysregulation, systemic inflammatory process, and development of ASD. Recent studies suggest that MI dysregulation plays an important role in the pathogenesis of inflammation, which may contribute to the manifestation of ASD symptoms [7].

Qiao et al., 2018, evaluated the salivary and dental microbiota of patients with ASD, identifying different characteristics of the

Received 12 March 2025; revised 2 April 2025; accepted 5 April 2025. Date of publication 15 April 2025; date of current version 1 May 2025. *(Corresponding author: Emilia Angela Lo Schiavo Arisawa.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the University of Vale do Paraíba, Brazil-CEP/UNIVAP under Application No. 1.749.222/2016.

Mayara Moniz Vieira Pinto and Emilia Angela Lo Schiavo Arisawa are with the Biostimulation and Tissue Repair Laboratory, University of Vale do Paraiba, São José dos Campos 12244-000, Brazil (e-mail: mirela@univap.br).

Leandro José Raniero is with the Nanosensors Laboratory, University of Vale do Paraiba, São José dos Campos 12244-000, Brazil.

Tanmoy Bhattacharjee is with the Faculty of Dentistry, University of Otago, Dunedin 9016, New Zealand.

This article has supplementary downloadable material available at https://doi.org/10.1109/JPHOT.2025.3561020, provided by the authors.

Digital Object Identifier 10.1109/JPHOT.2025.3561020

salivary and dental microbiota, when compared to neurotypical individuals [10]. Results obtained by saliva analysis include changes in salivary flow, associated with gender, which may imply in the protein concentration of this biofluid in the oral cavity of individuals with ASD, allowing its identification [11].

Several techniques are being explored to establish an easy, fast, and economical protocol for ASD diagnosis, including Fourier Transform Infrared Spectroscopy (FTIR). The technique gives the complete chemical profile of a sample, allowing qualitative and quantitative analysis of composition, concentration, and protein secondary structure changes. A significant advantage of this technique is its ability to identify unknown components in a mixture of different physical states of matter. It also allows fast, non-destructive analysis without using large sample amounts and reagents [12].

FTIR can be used to analyze various types of biological samples, for example, tissues, cells, etc. However, current research is focused on the use of biofluids such as blood, urine, and saliva as samples, owing to their ease of collection. Ildiz et al. have used blood as a sample for FTIR to ASD diagnosis [13]. However, blood collection is invasive and can be distressing to ASD children, hence urine and saliva need to be considered. Sarigul et al. have shown that urine-based FTIR can be used for ASD diagnostics [14]. However, there are no saliva-based FTIR studies for diagnosis of ASD.

Saliva collection is easy, fast, low-cost, and non-invasive techniques. Additionally, saliva collection generates less stress in patients diagnosed with ASD [15]. A considerable portion of saliva proteins comes from the bloodstream, entering by diffusion, filtration, or active transport by the gingival crevicular fluid [16], [17], thus allowing analysis of proteins from the entire vascular system. Saliva-based diagnostic techniques have been used to detect biomarkers associated with numerous systemic diseases [18]. Hence, the use of saliva for ASD diagnostics using FTIR must be explored and compared to previous studies.

Ease of use of FTIR is enhanced by multivariate analysis and machine learning models, hence we have explored several machine learning models to identify the best strategy for saliva FTIR based ASD diagnostics.

A. Methodology

1) Subjects: This study was approved by the Research Ethics Committee of the University of Vale do Paraíba, Brazil (CEP/UNIVAP- Opinion No. 1.749.222/2016). All parents and/or guardians signed the Free and Informed Consent Term (FICT), and the assessed individuals signed the Assent Term, developed in appropriate language for this audience and applied individually, both for neurotypical (Control) and for individuals with ASD. The criteria for inclusion of participants in this study were: age between 3 and 15 years, both sexes, previous diagnosis of ASD (study group) or neurotypical (control group), and absence of intraoral lesions. A total of 38 subjects were recruited for the study, 19 ASD and 19 Control.

2) Sample Collection: Before sample collection, the participants washed their mouths with distilled water 10 times. Unstimulated saliva was collected by the "spit" method in a 1.5 mL sterile tube, which was immediately labelled and refrigerated at 5 °C. Unstimulated saliva collection is best suited for sampling from ASD children [19], as stimulated saliva collection procedures can be a great stressor for behaviorally and socially challenged ASD children. The collected biofluid was subjected to centrifugation at 5.000 rpm for 10 min in the CF 16RN-Hitachi equipment to remove possible food residues. The supernatant was then transferred to a new sterile 1.5 mL microtube and frozen at −20 °C.

3) Analysis of Saliva by FTIR: Spectra acquisition procedure from saliva has been described earlier [20]. Briefly, samples were thawed passively at room temperature, homogenized on a vortex mixer, 20 μL deposited on calcium fluoride (CaF_2) windows, and dehydrated for 10 min in Eppendorf 5301 Concentrator. Spectra were acquired from 4-8 random points using a Spectrum 400 spectrophotometer, coupled to a microscope (Perkin-Elmer, Spotlight 400) and controlled by the Spotlight 400 Software (2H Flowchart in the 4000 - 900 cm^{-1} range at the resolution of 4 cm^{-1}, 32 scans being taken per spectrum).

4) Statistical Analysis: Baseline correction was performed using the OPUS software (Version 4.2, Bruker) by selecting the "rubberband" correction option and using 64 baseline points, followed by normalization between 0 and 1 using the same software. Mean spectra were then calculated using the baseline corrected and normalized spectra. Spectral deconvolution and peak fitting were performed using OriginPro 8.5 as described earlier [21]. For multivariate statistical and Machine Learning (ML) analysis, the spectra were preprocessed using established methods, namely first derivatization, spectral range selection, and area normalization [20], [21]. The spectral region 1250-1800 cm^{-1} was subjected to Principal Component Analysis (PCA), PC-Linear Discriminant Analysis (LDA), generalized additive model (GAM), k nearest neighbor (kNN), Naïve Bayes Method (NBM), Neural Network (NN), Support Vector Machine (SVM), and decision trees (Tree and Tree Bagger). The SVM training model was optimized and cross validated by Leave One Out Cross Validation (LOOCV). Sensitivity, precision, and specificity/recall classification parameters were calculated using formulas TP/(TP+FN), TP/(TP+FP), and TN/ (TN + FP), respectively, where TP, FN, FP, and TN are true positive, false negative, false positive, and true negative, respectively. Shapley values were calculated for each feature to quantify the impact of the different features on the SVM model, providing insight into SVM classifier interpretability. The ML analysis was performed using MATLAB 2024b (MathWorks, Natick, Massachusetts).

B. Results

Nineteen participants (2 females, 17 males, age-range 9.6 ± 2.5 years) previously diagnosed with ASD by qualified professionals and nineteen Controls (7 females, 12 males, age-range 12.2 ± 2.4 years) participated in this study. The shortest period for the onset of ASD symptoms reported was between 0 and 6 months, and the longest period was over 36 months; 9 participants among them showed symptoms between 12 and 18 months of age. ASD was diagnosed between 18 months and 9 years of age, with the mean age with SD of 5.3 ± 2.2 years. The first intervention for disease management was reported to have taken place between 2.5 and 7 years of age, with a mean age and SD of 4.4 ± 1.4 years.

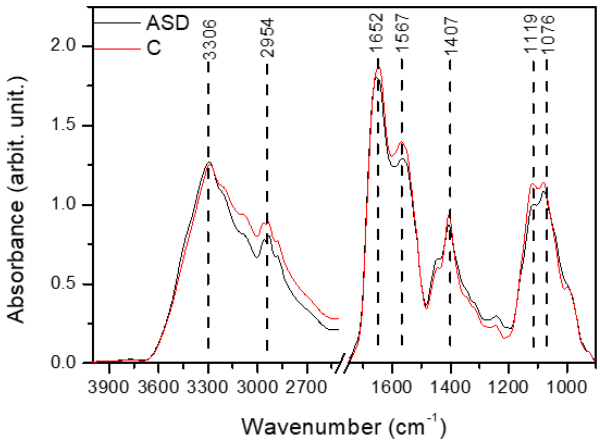

Fig. 1. Graph of mean FTIR spectra of saliva samples, from groups control (C) and autism spectrum disorder (ASD).

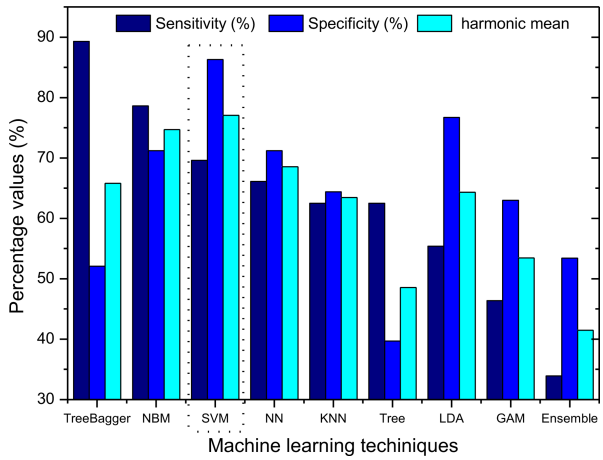

Fig. 2. Outcomes of machine learning techniques concerning levels of specificity and sensitivity. The harmonic means reveal the best result.

Mean FTIR spectra of saliva samples collected from above ASD and C participants (Fig. 1) showed 1076 (phosphate/DNA), 1119 (phosphate), 1407 (proteins), 1567 (nucleic acids), 1652 (Amide I), 2954 (lipids), and 3306 cm^{-1} (lipids) peaks [22], with visually the only difference between ASD and C being the 1119 cm^{-1} peak. Spectral deconvolution analysis found several new peaks, whose percentage areas under the curves for both ASD and C are presented in Supplementary Table I. Five bands that were found in ASD and not in C are 1521 cm^{-1} (nucleic acids), 1570 cm^{-1} (Amide II), 1599 cm^{-1} (Amide I), 1684 cm^{-1} (Amide I random coils), and 1735 cm^{-1} (lipids), while the bands 1493 (C-H bending), 1515 (Amide II) and 1582 cm^{-1} (phenyl ring) were found only in the spectra of the C group. The difference in the area under the curves for bands that are present in both groups show that the $\sim$ 1670 cm^{-1} band corresponding to Amide I of α-helix structures is considerably larger in ASD compared to C, while the $\sim$1635 cm^{-1} band corresponding to Amide I of β-pleated sheet structures is notably higher in group C. Lipids and carbohydrates also appear to be higher in ASD compared to C, while there is an apparent decrease in protein signatures in ASD compared to C. Mitochondrial dysfunction, oxidative stress, impaired methylation, and altered amino acid metabolism are the main biochemical mechanisms proposed to be underlying ASD. Inherited metabolic disorders strongly associated with ASD have been mapped to amino acid, carbohydrate, and fatty acid metabolism [23], amongst others. Hence, these biochemicals were chosen for area under the curve calculations in this study.

The next step was to test the capability of multiple machine learning methods - generalized additive model (GAM), k nearest neighbor (kNN), Naïve Bayes Method (NBM), Neural Network (NN), Support Vector Machine (SVM), and decision trees (Tree and Tree Bagger), to classify ASD spectra from Control. Results suggest that SVM gives the best classification (Fig. 2). The confusion matrix of the LOOCV SVM optimized model shows that 113 out of 122 ASD spectra were correctly predicted as ASD, and 146 out of 153 C spectra were correctly predicted as C. The sensitivity, precision, and specificity/recall values were 92, 94, and 95%, respectively.

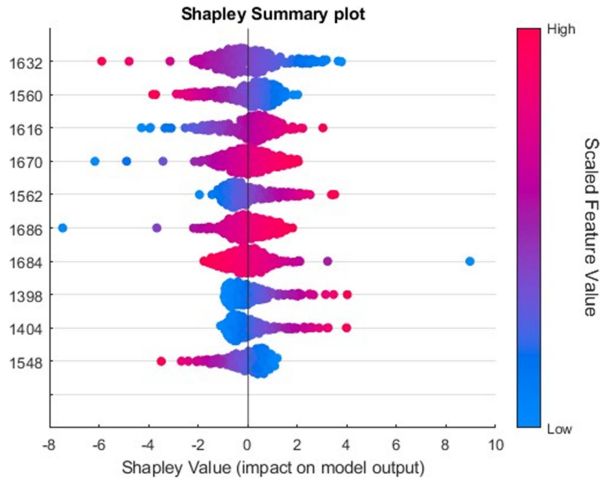

Fig. 3. Shapley values of the most important features that contribute to the classification of ASD and C. 1632 - β-sheets (Amide I), 1560 - Protein (Amide II), 1616 - intermolecular β-sheets, 1670 - β-turns (Amide I), 1562 - Protein (Amide II), 1686 - β-turns (Amide I), 1684 - β-turns (Amide I), 1398, 1404 – Protein, 1548 - Protein (Amide II).

To uncover the features responsible for the above classification, absolute Shapley values associated with each feature were averaged and ordered such that the most important features appear at the top. This is presented as a Shapley Summary plot (Fig. 3). In the case of this study, higher negative Shapley values suggest that the feature drives classification towards ASD, while higher positive Shapley values drive classification towards C. The summary plot shows that 1632, 1560, and 1548 cm^{-1} (β-sheets and protein signatures) drive classification towards ASD while 1616, 1562, 1398, and 1404 cm^{-1} (intermolecular β-sheets and protein signatures) drive classification towards C. β-turn signatures appear to drive classification in both directions: 1670 and 1686 towards C and 1684 cm^{-1} towards ASD.

C. Discussion

The present study explores the possibility of using saliva as a sample for Fourier Transform Infrared spectroscopy (FTIR) based autism diagnostics. The Optimized Support Vector

Machine (SVM) classifier could classify ASD and C with 92%, 94%, and 95% sensitivity, precision, and specificity, respectively, after cross-validation.

Ildiz et al. analyzed blood serum samples from 30 ASD and 30 control subjects with FTIR spectroscopy reporting 100% prediction accuracy using Principal Component Analysis (PCA) [13]. Sarigul et al. used urine samples with FTIR spectroscopy from 26 ASD and 26 control subjects [14]. PCA could not separate the two groups; there was a moderate level of overlap. The group did not explore supervised multivariate analysis for further classification. In our study, using saliva from 19 ASD and 19 control subjects as FTIR samples, PCA could not separate the two groups and there was considerable overlap (Supplementary Fig. 1).

Ildiz et al. study indicated that the blood serum of ASD patients has an increase in protein total contents, and a slight decrease in tyrosine compared to the control group, while the lipids total amount seems to be nearly identical [13]. Sarigul et al. study suggested lower-than-normal levels of uric acid, ammonium, and phosphate groups in the urine of ASD compared to control [14]. Our study found changes in saliva protein secondary structure content to differ between ASD and Control. Spectral analysis showed α-helix structures are considerably higher in ASD compared to C (Table I). Shapley values calculated to explain SVM classification suggest a change in β-sheets and β-turns, with β-sheets driving classification towards ASD, intermolecular β-sheets towards C, and β-turns towards ASD and C. β-strands are present in proteins that increase as well as decrease in ASD subject saliva compared to C. Changes in concentration of salivary proteins in ASD subjects compared to Control have been reported by several studies [24].

The results of the present study suggest that FT-IR spectroscopy is a promising tool for the diagnosis of autism spectrum disorder (ASD) using saliva samples, allowing to easily and quickly identify alterations in the composition of this fluid, mainly proteins, among diagnosed individuals with ASD, which could serve as objective, effective, and specific biomarkers for early diagnosis of this disorder. The technique can be a suitable tool for mass-screening. However, ensuring saliva is collected correctly without the presence of food particles or other confounding factors may be challenging for non-experts. The requirement for trained nurses for saliva collection can hinder mass-screening due to a shortage of staff or inability to access remote areas. Hiring expert staff will also increase the expense of mass-screening. Proper training and education among parents may mitigate this problem. Further studies can shed more light on this matter. Hence, a large-scale trial is thus warranted.

References

[1] R. P. Thom et al., "Neurodevelopmental disorders: Autism Spectrum disorder," in *Tasman's Psychiatry*, A. Tasman et al., Cham, Switzerland: Springer, 2024, doi: 10.1007/978-3-030-51366-5.

[2] "Data and statistics on autism spectrum disorder," CDC.gov, 2025. Accessed: Mar. 11, 2025 . [Online]. Available: https://www.cdc.gov/autism/data-research/index.html

[3] M. J. Maenner et al., "Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR Surveill. Summaries*, vol. 69, no. 4, pp. 1–12, Mar. 2020, doi: 10.15585/mmwr.ss6904a1.

[4] H. Brentani et al., "Autism spectrum disorders: An overview on diagnosis and treatment," *Braz. J. Psychiatry*, vol. 35, no. 1, pp. S62–S72, 2013, doi: 10.1590/1516-4446-2013-S104.

[5] C. Okoye et al., "Early diagnosis of Autism spectrum disorder: A review and analysis of the risks and benefits," *Cureus*, vol. 15, no. 8, 2023, Art. no. e43226, doi: 10.7759/cureus.43226.

[6] M. B. First, "Diagnostic and statistical manual of mental disorders, 5th edition, and clinical utility," *J. Nervous Ment. Dis.*, vol. 201, no. 9, pp. 727–729, Sep. 2013, doi: 10.1097/NMD.0b013e3182a2168a.

[7] X. Kong et al., "New and preliminary evidence on altered oral and gut microbiota in individuals with autism spectrum disorder (ASD): Implications for ASD diagnosis and subtyping based on microbial biomarkers," *Nutrients*, vol. 11, no. 9, 2019, Art. no. 2128.

[8] B. D. Needham, W. Tang, and W.-L. Wu, "Searching for the gut microbial contributing factors to social behavior in rodent models of autism spectrum disorder," *Devlop. Neurobiol.*, vol. 78, no. 5, pp. 474–499, Feb. 2018.

[9] D.-W. Kang et al., "Long-term benefit of Microbiota transfer therapy on autism symptoms and gut microbiota," *Sci. Reports*, vol. 9, 2019, Art. no. 5821.

[10] Y. Qiao et al., "Alterations of oral microbiota distinguish children with autism spectrum disorders from healthy controls," *Sci. Reports*, vol. 8, 2018, Art. no. 1597.

[11] M. M. Souza et al., "Salivary flow and salivary proteins in male and female children with autism spectrum disorder: Pilot study," *Revista Odontologia UNESP*, vol. 51, 2022, Art. no. e20220041.

[12] P. C. Caetano Jr., J. F. Strixino, and L. Raniero, "Analysis of saliva by Fourier transform infrared spectroscopy for diagnosis of physiological stress in athletes," *Res. Biomed. Eng.*, vol. 31, no. 2, pp. 116–124, 2015, doi: 10.1590/2446-4740.0664.

[13] G. O. Ildiz, S. Bayari, A. Karadag, E. Kaygisiz, and R. Fausto, "Fourier transform infrared spectroscopy based complementary diagnosis tool for Autism spectrum disorder in children and adolescents," *Molecules*, vol. 25, no. 9, Apr. 2020, Art. no. 2079, doi: 10.3390/molecules25092079.

[14] N. Sarigul, L. Bozatli, I. Kurultak, and F. Korkmaz, "Using urine FTIR spectra to screen autism spectrum disorder," *Sci. Reports*, vol. 13, no. 1, Nov. 2023, Art. no. 19466, doi: 10.1038/s41598-023-46507-z.

[15] S. K. Putnam et al., "Comparison of saliva collection methods in children with high-functioning autism spectrum disorders: Acceptability and recovery of cortisol," *Child Psychiatry Hum. Develop.*, vol. 43, no. 4, pp. 560–573, Aug. 2012, doi: 10.1007/s10578-012-0284-3.

[16] J. Marchezan, E. G. A. W. Dos Santos, I. Deckmann, and R. D. S. Riesgo, "Immunological dysfunction in autism spectrum disorder: A potential target for therapy," *Neuroimmunomodulation*, vol. 25, no. 5/6, pp. 300–319, 2018, doi: 10.1159/000492225.

[17] J. M. Yoshizawa, C. A. Schafer, J. J. Schafer, J. J. Farrell, B. J. Paster, and D. T. Wong, "Salivary biomarkers: Toward future clinical and diagnostic utilities," *Clin. Microbiol. Rev.*, vol. 26, no. 4, pp. 781–791, Oct. 2013, doi: 10.1128/CMR.00021-13.

[18] C.-Z. Zhang et al., "Saliva in the diagnosis of diseases," *Int. J. Oral Sci.*, vol. 8, no. 3, pp. 133–137, Sep. 2016, doi: 10.1038/ijos.2016.38.

[19] A. Galiana-Simal, V. Muñoz-Martinez, P. Calero-Bueno, M. Vela-Romero, and L. Beato-Fernandez, "Towards a future molecular diagnosis of autism: Recent advances in biomarkers research from saliva samples," *Int. J. Dev. Neurosci.*, vol. 67, no. 1, pp. 1–5, Jun. 2018, doi: 10.1016/j.ijdevneu.2018.03.004.

[20] D. R. Paschotto, B. Pupin, T. T. Bhattacharjee, and L. E. S. Soares, "Saliva preparation method exploration for ATR-FTIR spectroscopy: Towards biofluid based disease diagnosis," *Anal. Sci.*, vol. 36, no. 9, pp. 1059–1064, Sep. 2020, doi: 10.2116/analsci.20P029.

[21] T. T. Bhattacharjee, M. C. Nicodemo, L. B. Sant'Anna, E. A. Lo Schiavo Arisawa, and L. Raniero, "Tendinopathy diagnosis and treatment monitoring using attenuated total reflectance-fourier transform infrared spectroscopy," *J. Biophotonics*, vol. 11, no. 4, Apr. 2018, Art. no. e201700256, doi: 10.1002/jbio.201700256.

[22] Z. Movasaghi, S. Rehman, and D. I. ur Rehman, "Fourier transform infrared (FTIR) spectroscopy of biological tissues," *Appl. Spectrosc. Rev.*, vol. 43, no. 2, pp. 134–179, 2008, doi: 10.1080/05704920701829043.

[23] U. D. Senarathne et al., "Biochemical, genetic and clinical diagnostic approaches to autism-associated inherited metabolic disorders," *Genes*, vol. 14, no. 4, 2023, Art. no. 803, doi: 10.3390/genes14040803.

[24] A. G. Ngounou Wetie, K. L. Wormwood, S. Russell, J. P. Ryan, C. C. Darie, and A. G. Woods, "A pilot proteomic analysis of salivary biomarkers in autism spectrum disorder," *Autism Res.*, vol. 8, no. 3, pp. 338–350, Jun. 2015, doi: 10.1002/aur.1450.