A Wearable Device Dataset for Mental Health Assessment Using Laser Doppler Flowmetry and Fluorescence Spectroscopy Sensors

BLIND

Blind Blind Blind

Abstract—In this study, we present a large, novel dataset collected using a non-invasive wearable device equipped with Laser Doppler Flowmetry (LDF) and Fluorescence Spectroscopy (FS) sensors to predict depression, anxiety, and stress levels via the DAS-21 questionnaire. To our best knowledge, this is the largest and the most generalized dataset ever collected for both LDF and FS studies. The device captures cutaneous blood microcirculation parameters, and wavelet analysis of the LDF signal extracts key rhythmic oscillations. The dataset, collected from 132 volunteers aged 18-94 from 19 countries, explores relationships between physiological features, demographics, lifestyle habits, and health conditions. We employ a variety of machine learning methods to classify stress detection, in which LightGBM is identified as the most effective model for stress detection, achieving a ROC AUC of 0.7168 and a PR AUC of 0.8852. In addition, we also incorporated Explainable Artificial Intelligence (XAI) techniques into our analysis to investigate deeper insights into the model's predictions. Our results suggest that younger individuals and those with a higher Body Mass Index (BMI) have a greater likelihood of experiencing mental health conditions like stress and anxiety. This study paves the way for using LDF/FS wearable devices for comprehensive health assessment and highlights the importance of interpretable models in mental and cardiovascular health monitoring.

Index Terms—Laser doppler flowmetry (LDF), Fluorescence spectroscopy (FS), Smart health applications, Explainable AI, Machine learning

I. INTRODUCTION

A. Motivation and Literature Review

Over the past two decades, global incidences of Common Mental Disorders (CMDs), particularly anxiety and depression, have fluctuated significantly and increased substantially due to improved awareness and diagnosis in healthcare settings [1]. However, the increase in CMDs is not uniform across age groups, with higher rates among younger individuals due to changing social pressures and lifestyle factors [2]. Economic conditions and public health crises also influence mental health trends, highlighting the need for adaptable and accessible mental health services in the healthcare system. [3]. Mental health has gained significant attention, particularly after the COVID-19 pandemic, which exacerbated mental health issues [4], [5].

In the United Kingdom, over 25% of individuals experience a mental health disorder annually, with 1 in 6 adults facing anxiety or depression weekly [1]; stress leading to overeating (46%), increased alcohol consumption (29%), and elevated smoking rates (16%)[2]. CMDs harm various body systems, including raising blood pressure and heart risks in the cardiovascular system, impairing learning and mood in the nervous system, causing tension and fatigue in muscles, resulting in shallow breathing, and leading to weight changes and diabetes risk in metabolism. Ultimately, stress extensively affects both mental and physical well-being [6].

Stress can have a detrimental impact on various body systems [7]. Prolonged stress can elevate blood pressure and heart rate, increasing the risk of cardiovascular diseases [8]. It also affects the nervous system, leading to cognitive decline, mood disorders, and an increased risk of mental disorders [9]. Muscular tension, soreness, and fatigue can result from stress, impairing daily activities [10]. Changes in breathing patterns due to stress can lead to respiratory issues [11]. Additionally, stress disrupts metabolism, potentially causing weight changes and increasing the risk of diabetes [12]. In conclusion, stress negatively affects both mental and physical health, impacting systems such as cardiovascular, nervous, muscular, respiratory, and metabolic.

Mental health assessment encompasses various methods to ensure a comprehensive and accurate understanding. Standardized tests like DAS (Depression Anxiety Stress Scales) [13], the Beck Depression Inventory (BDI) [14], and the Beck Anxiety Inventory (BAI) [15] Clinical interviews, in structured, semi-structured, or unstructured formats, measure levels of depression, anxiety, and stress, helping psychologists gather detailed information through specific questions and conversations. Biological assessments, including tests for neurotransmitter levels like serotonin and dopamine, and electroencephalograms (EEGs) to monitor brain activity, also play a crucial role [16], and functional magnetic resonance imaging (fMRI) to observe brain activity during psychological tasks [17]. Biosensors for psychiatric biomarkers (e.g., cortisol, dopamine, serotonin) can diagnose and manage disorders via samples from blood, saliva, urine, and sweat. They offer high sensitivity, selectivity, and real-time monitoring, but face challenges like environmental accuracy, high costs, and data integration. Therefore, further development is needed for better effectiveness [18].

The DAS-21 questionnaire, a short version of the 42-

---

[1] https://www.mind.org.uk/news-campaigns/news/mind-urges-the-nation-speak-to-us-during-mental-health-awareness-week/

[2] https://www.myndup.com/blog/mental-health-statistics-2023

item DAS, includes 21 items divided into three subscales: Depression, Anxiety, and Stress. It assesses motivation loss, anxiety symptoms, and irritability, respectively. Validated in clinical and community settings, the DAS-21 shows excellent internal consistency with Cronbach's alpha values of 0.94 for depression, 0.87 for anxiety, and 0.91 for stress. The DAS-21 severity levels and cutoff points classify and promptly support patients [19]. Intense emotions like anxiety or anger can affect the hands by altering blood flow and muscular electrical activity, causing muscle tension or relaxation [20]. Despite many articles on blood circulation in such individuals, none compare blood circulation variability in stressed vs. non-stressed people. This study demonstrated the wearable device's ability to differentiate cardiovascular parameters between stress and non-stress groups on both middle fingers.

Professor E. Rafailov's research group at Aston University has developed LDF/FS wearable devices using VCSEs, showing comparable signal responses to conventional monitors in volunteer assessments [21]. These devices employ Laser Doppler Flowmetry (LDF) and fluorescence spectroscopy (FS) for non-invasive early detection of vascular complications in diabetes and other conditions. LDF assesses tissue perfusion, oxygen saturation, and blood volume, while FS detects metabolic activity changes and AGEs accumulation, contributing to microvascular damage and inflammation in diabetes.

LDF is a non-invasive method for estimating perfusion in the microcirculation [22]. Introduced over 30 years ago, the technique uses laser radiation to probe tissue and analyze backscatter from moving red blood cells, primarily Hemoglobin (Hb). The main parameter recorded is the microcirculation or perfusion index, essential for organ nutrition, adaptation, and regulation. The method uses wavelet transformation, specifically adaptive wavelet analysis with complex-valued Morlet wavelets, to assess microvessel oscillatory processes over a wide frequency range. This has been the standard for over 15 years, replacing Fast Fourier Transform (FFT) and Butterworth filters [23]. Continuous wavelet transformation is preferred for non-stationary LDF-gram (perfusion) due to its optimal "time-frequency" resolution, effectively tracking frequency and amplitude fluctuations in blood flow signals [24]. The FS method uses laser probing to record fluorescence spectra of metabolic coenzymes, measuring NADH and FAD fluorescence intensity. This detects changes in metabolic activity in endothelial cells, indicating various physiological and pathological processes, and identifying cellular metabolic disorders related to diseases [25].

Several studies have utilized wearable devices to assess blood microcirculation across diverse patient groups. Older adults typically exhibit higher perfusion levels in areas like the middle palm and dorsal forearm due to thinner skin, aiding in diagnostic precision [24]. Conversely, younger individuals often show elevated wavelet parameters in blood perfusion oscillations, suggesting broad applicability in various pathologies. In endocrinology, wireless LDF devices have been used to evaluate microcirculatory function in type 2 diabetes patients and healthy individuals across different age brackets, revealing significant variations in perfusion levels [26]. Notably, studies monitoring diabetes patients receiving intravenous alpha-lipoic acid therapy have shown improvements in microcirculatory and nutritional blood flow, particularly in limbs affected by diabetic complications [27]. Additionally, wearable LDF devices have been instrumental in diagnosing vascular disorders during COVID-19 recovery, highlighting disruptions in microcirculatory function [28].

Wearable devices with LDF and FS channels offer a promising approach for assessing microcirculation and obtaining comprehensive physiological and metabolic information. While these studies demonstrate their potential under normal and pathological conditions, further research with larger cohorts is essential for clinical implementation. One of the crucial tasks is to investigate the effects of various treatment protocols and lifestyle changes on microcirculatory and metabolic parameters using these wearable devices. Another important direction is to develop machine learning algorithms for automated data analysis and interpretation, which can significantly enhance the diagnostic capabilities of wearable devices. Our research focuses on building a diverse dataset for mental health detection using a non-invasive wearable device equipped with LDF and FS channels. By exploring subcutaneous blood microcirculation across demographics, we aim to provide valuable insights and pioneer the development of a large dataset for mental health assessment.

B. Contribution

In this study, we make three key contributions to the field of mental health assessment, placing particular emphasis on our data collection methods and the application of Explainable AI (XAI):

- **Establishing the largest and the most generalized dataset ever collected for both LDF and FS studies**: We address the need for robust datasets in the field by creating a novel data repository comprised of physiological signals captured using wearable devices. The dataset contains 132 patients, which is specifically chosen for its relevance to mental health and is further enriched by integrating self-reported DAS scores obtained through the validated depression, anxiety, and stress scale-21 (DAS-21) questionnaire.
- **Exploring numerous machine-learning algorithms for DAS prediction**: We move beyond traditional approaches that solely focus on achieving high prediction accuracy. We delve into the feasibility of utilizing various machine learning algorithms for predicting DAS levels.
- **Unveiling the *"AI black box"* by using XAI**: Recognizing the critical role of interpretability in mental health applications, we employ XAI techniques to investigate the decision-making behind a machine learning model. By employing XAI, we aim to illuminate the specific features within the wearable device data that exert the strongest influence on the health issues prediction of a person.

II. STUDY DESIGN AND DATASET DESCRIPTION

STEP 1: Screening
• Recruit and provide the participant information sheet (PIS) outlining the study details for the volunteer
• Get basic information and consent capture.

STEP 3: Collect personal data
• Check high and weigh
• Measure blood pressure
• Complete a questionnaire including DASS-21

STEP 2: Collect real-time sensor data
• Preparation: Lie down comfortably on the bed
• Measure blood circulation for 15 minutes with LDF and FS devices

STEP 4 – Repeat data collection
• Volunteers are asked to attend 30-minute sessions for 5 days across two weeks.
• Time 11:00 am and 6:00 pm on each of the 5 selected days

Fig. 1: Data collection workflow.

There are four steps in data collection as shown in Fig. 1. Firstly, participants were recruited from the general population and included volunteers aged 18 and above. To ensure accurate blood perfusion measurements, individuals with any dermatological conditions on both hands and middle fingers were excluded from the study. Before commencing the study, all participants were provided with a detailed explanation of the study design and its objectives. After giving informed consent, participants completed a questionnaire detailing their current health status, including medication history, alcohol consumption within the past 24 hours, and exercise habits such as cycling, treadmill, or jogging.

(a) A stress instance of data collected using the wearable devices. The subject is a 36-year-old female with moderate stress, anxiety, and depression (right hand).
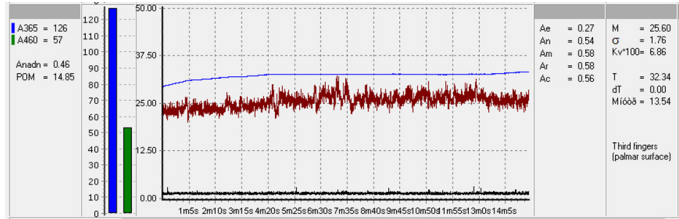
(b) An instance of well-being data collected using the wearable devices. The subject is a 27-year-old female (right hand).

Fig. 2: Data instances collected using the wearable devices: (a) stress instance, (b) well-being instance.

Sequentially, blood perfusion parameters were measured non-invasively with participants in a supine position to ensure physical and mental rest. To minimize external stimuli, participants were instructed to abstain from reading, writing, or talking during the test. Blood perfusion data were collected from sensors placed on the middle fingertips of both left and right hands for a duration of eight minutes. To control potential confounding factors, participants were asked to refrain from consuming caffeine and alcohol-containing drinks at least twelve hours before the designated measurement time.

Figure 2a shows the data measured from a stressed individual, with data from the left hand illustrated on the top and data from the right hand on the bottom. Similarly, Figure 2b presents an instance of well-being data collected using wearable devices. As observed, the data from the stressed individual exhibits significant fluctuations, while the data from the well-being individual is more stable. In addition, the definitions of the measurement device parameters are described following Table I.

TABLE I: Definitions of the measurement device parameters.

Parameters	Definition
M	Microcirculation index, indicating the average perfusion of microvessels (in PU).
σ	Mean square deviation of blood flow oscillation amplitude (in PU).
Kv	Coefficient of blood flow variability.
A365	Backscatter amplitude at the laser source wavelength for NADH excitation.
A460	NADH fluorescence amplitude at 460 nm.
NADH	Relative amplitude of NADH fluorescence, considering the optical characteristics of the study tissue region.
POM	Index of oxidative metabolism linked to the nutritional component of blood perfusion and NADH coenzyme fluorescence amplitude.
Ae	Average maximum amplitude of blood flow within the endothelial oscillation range.
An	Average maximum amplitude of blood flow within the neurogenic oscillation range.
Am	Average maximum amplitude of blood flow within the myogenic oscillation range.
Ar	Average maximum amplitude of blood flow within the respiratory oscillation range.
Ac	Average maximum amplitude of blood flow within the cardiac oscillation range.
Fe	Endothelial oscillation frequency (0.0095 - 0.02 Hz).
Fn	Neurogenic oscillation frequency (0.02 - 0.06 Hz).
Fm	Myogenic oscillation frequency (0.06 - 0.16 Hz).
Fr	Respiratory oscillation frequency (0.16 - 0.4 Hz).
Fc	Cardiac oscillation frequency (0.4 - 1.6 Hz).
T	Temperature at the measurement site.

Following the 15-minute blood circulation measurement, we measured height and weight. Next, the participants completed the DAS-21 questionnaire, which assesses how much each statement applied to them over the past week. After completing the questionnaire, their blood pressure was measured. The measurements were taken twice a day: in the morning (around 11.00, before lunch) and in the afternoon (around 15.00, after lunch) for any five days over two consecutive weeks.

The DAS-21 is used to assess key symptoms of depression, anxiety, and stress, as well as patient reactions to treatment. It has been proven to have adequate psychometric properties and

[3]https://anonymous.4open.science/r/LDF_private-1B8E

is equivalent to other accurate scales. The 21 items comprise three self-reported scales, each with seven elements graded on a Likert scale from 0 to 3. Depression, anxiety, and stress scores are measured by summing the scores of the related items. Since the DAS-21 is a shorter version of the original 42-item DAS, the score for each subscale must be multiplied by 2 to calculate the final score. Recommended cut-off scores for conventional severity labels (normal, moderate, severe) are calculated following Table II. Scores on the DAS-21 will need to be multiplied by 2 to calculate the final score.

TABLE II: Scores on the DAS-21 will need to be multiplied by two to calculate the final score.

Level	Depression	Anxiety	stress
Normal	0-9	0-7	0-14
Mild	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Severe	21-27	15-19	26-33
Extremely Severe	28+	20+	34+

According to the manual, the ratings are classified as: "normal, mild, moderate, severe, or extremely severe"; all those who exhibit any signs of stress, anxiety, or depression, we referred to as the well-being group, and the remaining individuals will be classified as the wellbeing group. This allowed for real-time control of the course of the experiment and analysis of the recorded parameters.

The displayed parameters show the raw data of blood perfusion, temperature, and the movement of the fingertip and wrist. After acquiring the data, the oscillation rhythms of each measurement were analyzed using the built-in module "wavelet analysis". This wavelet analysis determines the maximum amplitude of blood perfusion and corresponding data for each of the five oscillations: Five rhythmic oscillations are isolated from LDF recordings with the help of wavelet analysis; endothelial (frequency interval 0.0095–0.02 Hz), neurogenic (0.02–0.06 Hz), myogenic (0.06–0.16 Hz), respiratory (0.16–0.4 Hz), and cardiac or pulse rhythm (0.4–1.6 Hz).

As illustrated in Fig. 3, the total number of people with mental health issues reaches 27.3% of the population, with over 50% of them experiencing combined stress, anxiety, and depression. The incidence rates of stress, anxiety, and depression are 24.5%, 22%, and 18.2% respectively, mostly at mild levels, accounting for 17.2%, 13.6%, and 12.8% in these groups. The extremely severe level is highest in the anxiety group at 3.0%, while in the other two groups, it is below 1%.

A. Data Analysis

Similar to a medical record, the dataset includes essential demographic details for each participant. Our research involved a wide range of ages (18-94) with an average participant age of 40. We also ensured participant diversity by including a variety of races (Asian, White, African) and genders (55.2% male, 44.5% female). Most of the participants do not smoke cigarettes, and the rate of people used to smoke is small compared to smoked people and non-smoke people.
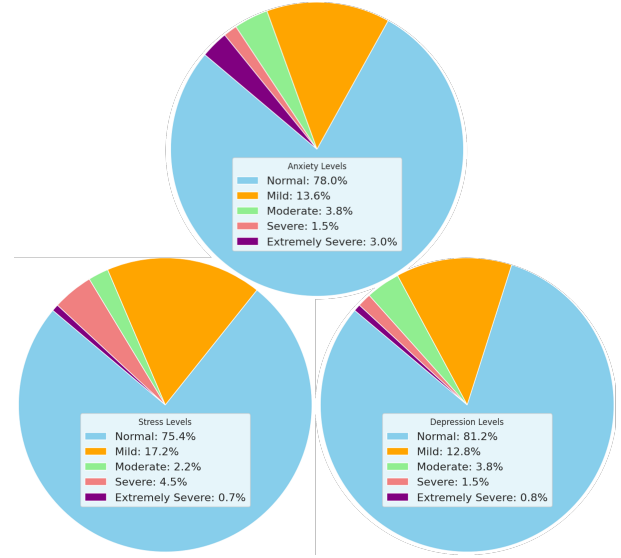

Fig. 3: Distribution of stress levels, anxiety level, and depression level.

Moreover, the number of people who have problems with health issues is significantly smaller than normal people.
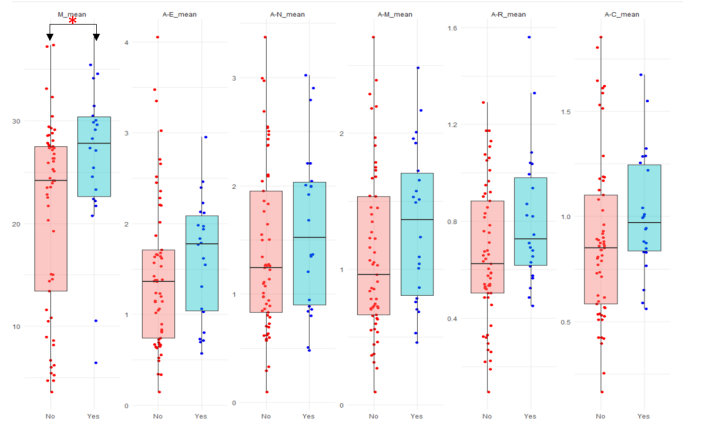

Fig. 4: Blood perfusion (M*) with standard deviation and Maximum amplitude with a standard deviation of the endothelial (A-E), neurogenic (A-N), myogenic (A-M), breath (A-R) and pulse (A-C) mechanism for Wellbeing vs Non-wellbeing, *, $p < 0.01$, Mann-Whitney U test.

The dataset incorporates data collected from wearable devices worn by participants. We even recorded details like sleep patterns (sleeping vs. awake) and the hand used for blood sample collection during the study. These devices continuously monitor various physiological and activity-related aspects, providing real-time health information. Examples of data collected include Body Mass Index (BMI), heart rate, and caloric expenditure. Each data point is linked to a specific time, which allows us to analyze trends and potential correlations between a participant's physiological responses and their mental state over time. The experiment duration is around 15 minutes, at the end of the experiment.

Fig. 7 indicate that non-wellbeing individuals exhibit higher perfusion parameters and amplitude variations compared to healthy individuals. For the perfusion parameter M, healthy individuals have a mean value of 21.02 (95% CI, 4.73 - 35.59), whereas non-wellbeing individuals have a significantly higher M value of 26.49 (95% CI, 8.59 - 36.71), (p=0.016, Mann-Whitney U test). The endothelial, neural, muscle, respiratory, and cardiovascular amplitude variations all tend to be higher by more than 0.2, but the differences are not statistically significant.

As shown in Table XX, non-wellbeing individuals exhibit a statistically significant higher amplitude of perfusion parameter fluctuations (δ) compared to their wellbeing counterparts, with values of 4.79 (95% CI, 2.25 - 7.65) versus 3.64 (95% CI, 0.98 - 6.93), (p=0.007, Mann-Whitney U test). Additionally, they have a significantly higher temperature at the measurement site, 33.21 (30.14 - 35.82) compared to 30.65 (95% CI, 22.68 - 35.6), (p=0.01, Mann-Whitney U test). Moreover, the metabolic index (POM) at the measurement site is also significantly elevated in non-wellbeing individuals, with values of 11.42 (95% CI, 1.95 - 24.42) versus 7.7 (95% CI, 0.85 - 20.43), (p=0.01, Mann-Whitney U test).

The DAS-21 questionnaire provides a standardized psychological assessment of each participant's mental well-being. This 21-question survey specifically focuses on measuring levels of depression, anxiety, and stress. Participants' responses are carefully recorded and scored based on established guidelines, generating quantifiable metrics for each dimension (depression, anxiety, stress). These scores offer valuable insights into potential mental health concerns. Additionally, timestamps associated with each DAS-21 completion allow us to track changes in a participant's mental state over time.

By combining personal information, DAS-21 scores, and wearable sensor data, we create a rich dataset that facilitates the development of accurate machine-learning models for predicting DAS levels.

III. MACHINE LEARNING AND EXPLAINABLE ARTIFICIAL INTELLIGENCE

A. Machine-learning Models for DAS Prediction

To identify the most effective approach for predicting depression, anxiety, and stress levels, we explored various machine learning algorithms including Support Vector Machine (SVM), Random Forest Classification, Gradient Boosting Classifier, CatBoost, LightGBM, as well as Multi-layer Perceptron (MLP) [29]. In addition, we employ two primary approaches to train machine learning models for predicting DAS levels: binary classification and multi-class classification. Both approaches leverage data from the DAS-21 questionnaire alongside potentially other features from the collected dataset. In addition, we consider three cases to investigate the models' performances: Using all collected features, using only features extracted from wearable devices, and using top-10 important features.

For binary classification, this approach simplifies the prediction task by transforming the DAS levels into a binary classification problem. We categorize participants into two classes based on their DAS-21 scores:

- Normal: This class comprises participants who score within the normal range for depression, anxiety, and stress according to established DAS-21 scoring guidelines.
- Abnormal: This class encompasses participants whose DAS-21 scores indicate potential symptoms of depression, anxiety, or stress.

For multi-class classification, this approach aims for a more granular prediction by treating DAS levels as a multi-class classification problem. Instead of collapsing mental health states into two categories, we define multiple classes based on the established DAS-21 scoring ranges: Normal, stress, stress Anxiety, and stress anxiety depression.

In machine learning, dividing the dataset into training and testing subsets is crucial for evaluating model performance. In our ablation study, we use three train-evaluate techniques: Split 80:20, patient-wise 5-folds (not sample-wise), and Leave-one-patient-out (LOPO) [30]. By doing this, we ensure that the model is evaluated on its ability to perform on new patients not seen during training.

To assess the performance of the machine learning models for predicting Depression, Anxiety, and stress (DAS) levels, we employ two key evaluation metrics: Receiver Operating Characteristic (ROC) AUC (Area Under the Curve) and Precision-Recall (PR) AUC. These metrics provide a comprehensive assessment of the model's discriminative ability and its performance in handling class imbalances.

B. Explainable AI

In healthcare applications, understanding the reasoning behind a model's predictions for DAS levels is crucial for building trust and confidence in its outputs. This empowers healthcare professionals and researchers to make informed decisions based on the predicted DAS levels and the underlying factors influencing those predictions. In this study, we leverage SHAP (Shapley Additive Explanations) to achieve interpretability and gain insights into the model's decision-making process for DAS prediction [31]. SHAP assigns an attribution value (SHAP value) to each feature for a given DAS prediction. High positive SHAP values indicate that the feature has a strong positive influence on the predicted DAS level (potentially indicating a higher likelihood of depression, anxiety, or stress). Conversely, low negative SHAP values signify a negative influence (indicating a lower likelihood). This interpretability allows us to answer several key questions:

- Identification of the key physiological and psychological indicators: What are the features from wearable sensor data and questionnaire scores of a patient that have the most significant influence on the model's predictions?
- Validation of model fairness and mitigation of mias: Are the model's predictions fair across different demographics (age, gender, etc.)? Examining SHAP values across these groups helps ensure that the model is not unfairly biased toward certain populations.

TABLE III: The results of parameters with standard deviation for Wellbeing vs Non-wellbeing: Kv100, δ*, T*, A365, A460, Anadn, POM*, F-E; F-N; F-M; F-R; F-C, *, p<0.01, Mann-Whitney U test.

Group	Kv100 mean	δ mean	p-value	T mean	p-value	A365 mean	A460 mean	Anadn mean	POM mean	p-value
All	21.09 (6.86 - 49.55)	3.96 (1.21 - 7.41)		31.36 (22.95 - 35.79)		86.82 (4.42 - 158.6)	59.3 (12.92 - 106.52)	1.01 (0.4 - 4.54)	8.74 (0.99 - 22.15)	
Wellbeing	21.12 (6.71 - 48.75)	3.64 (0.98 - 6.93)	0.0073	30.65 (22.68 - 35.6)	0.0184	85.43 (9.5 - 130.9)	60.64 (17.2 - 106.8)	1.01 (0.41 - 4.77)	7.7 (0.85 - 20.43)	0.0106
Non-Wellbeing	21 (7.56 - 48.44)	4.79 (2.25 - 7.65)		33.21 (30.14 - 35.82)		90.43 (2.52 - 159.81)	55.81 (12.05 - 88.06)	1 (0.39 - 4.16)	11.42 (1.95 - 24.42)	

- Enhanced model transparency: How does the model arrive at its predictions? By explaining the rationale behind the model's predictions through SHAP values, we can foster trust and confidence in its use among healthcare professionals and researchers.

IV. EXPERIMENTAL RESULTS

A. All Features with 80:20 Split

In this section, we present the results of our investigation into using machine learning models to predict stress levels based on data from the DAS-21 questionnaire and potentially other features within our dataset. We employed both binary and multi-class classification approaches, evaluating the models on a random 80/20 train-test split to ensure generalizability.

1) Binary Classification: Our initial focus was on a binary classification task, aiming to identify individuals with potential mental health concerns based on their DAS-21 scores. For binary classification, the performance of the models on binary classification tasks is summarized in Table IV.

From the table, LightGBM emerged as the best-performing model, achieving the highest ROC AUC of 0.9941 and PR AUC of 0.9982. Gradient Boosting and MLP also demonstrated strong performance, with ROC AUC values of 0.9751 and 0.9322, respectively. In contrast, Catboost and Random Forest showed relatively lower performance, indicating that they might not be as effective for this particular binary classification task.

TABLE IV: Performance for different models: All features with 80:20 split, binary classification

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest	MLP
ROC AUC	0.9751	0.7320	0.9941	0.9199	0.8145	0.9322
PR AUC	0.9911	0.9104	0.9982	0.9720	0.9330	0.9767

2) Multi-class Classification: In addition to predicting whether a person has a mental issue or not, we also explored a multi-class classification task, aiming to predict not only the presence of stress but also its severity level. In particular, Table V details the performance metrics of the models on multi-class classification tasks, with the notable absence of MLP results.

LightGBM again stands out, achieving near-perfect Macro ROC AUC scores and high precision, recall, and F1 scores. Gradient Boosting and SVM also performed well, with Gradient Boosting showing a balanced performance across all metrics. Catboost and Random Forest had lower scores, suggesting limitations in handling the complexities of multi-class classification in this context.

TABLE V: Performance for different models: All features with 80:20 split, multi-class classification

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest
Macro ROC AUC One-vs-Rest	0.8043	0.6932	0.9962	0.973	0.8695
Macro ROC AUC One-vs-One	0.8302	0.6875	0.993	0.9574	0.7952
Macro Precision	0.6238	0.1723	0.9875	0.4417	0.2966
Recall	0.5319	0.2108	0.9085	0.3799	0.1845
F1-score	0.5152	0.1808	0.9391	0.375	0.1783

TABLE VI: Top 10 important features using Gradient Boosting, Catboost, and LightGBM when conducting binary prediction with an 80:20 split. The meaning of each feature is explained in Table I.

Order	Gradient Boosting		Catboost		LightGBM	
	Feature	Importance	Feature	Importance	Feature	Importance
1	BMI_index	0.279819	Age	55.244066	BMI_index	22
2	Heart Rate	0.163589	Type of skins	29.933423	Heart Rate	13
3	Age	0.160837	Weight	11.040225	Age	13
4	Type of skins	0.156214	δ	3.782286	Weight	9
5	Weight	0.097077	Type of data	0.000000	Height	8
6	T	0.050864	F_Ae	0.000000	M	6
7	Height	0.044463	Level of BP	0.000000	T	6
8	A460	0.011797	Smoking routine	0.000000	A460	5
9	Anadn	0.009720	BMI_index	0.000000	Kv100	2
10	M	0.009113	Height	0.000000	Type of skins	2

3) Feature Importance: To understand the factors influencing the models' predictions, we analyzed the importance of various features. Feature importance was assessed using Gradient Boosting, Catboost, and LightGBM models, as summarized in Table VI and VII. The tables highlight the top 10 important features identified by each model. In both tables, features such as heart rate, BMI, weight, T (temperature), and type of skin consistently rank high in the top ten importance for most models. This suggests that physiological factors significantly influence the models' stress predictions. Other features including age, POM, A365, and Anadn also appear to be relevant to some degree, depending on the model.

B. All Features with Cross-Validation

In the field of health issue analysis, ensuring the robustness and reliability of predictive models is paramount. To achieve this, we employ cross-validation techniques such as k-fold cross-validation and LOPO cross-validation.

1) Binary Classification with LOPO: LOPO cross-validation is particularly relevant in medical studies, where

TABLE VII: Top 10 important features using Gradient Boosting, Catboost, and LightGBM for multi-class classification with an 80:20 split. The meaning of each feature is explained in Table I.

Order	Gradient Boosting		Catboost		LightGBM	
	Feature	Importance	Feature	Importance	Feature	Importance
1	Heart Rate	0.682942	Heart Rate	87.150766	Weight	18
2	A365	0.179464	Type of skins	8.122651	Height	9
3	BMI_index	0.089107	Anadn	3.149739	BMI_index	8
4	Type of skins	0.043119	δ	1.576843	Heart Rate	7
5	Height	0.004020	F_An	0.000000	Type of skins	7
6	Age	0.001002	Level of BP	0.000000	A365	7
7	POM	0.000173	Smoking routine	0.000000	Age	5
8	T	0.000173	BMI_index	0.000000	POM	5
9	F_An	0.000000	Height	0.000000	T	5
10	Level of BP	0.000000	Weight	0.000000	F_Ar	2

patient-specific variations can significantly impact the model's predictions. Table VIII presents the performance metrics for various machine learning models when evaluated using the LOPO cross-validation method for binary classification. LOPO is a stringent evaluation method where the model is trained on all patients except one, who is then used as the test set. This process is repeated for each patient, ensuring that the model's performance is tested on unseen data in each iteration.

TABLE VIII: Performance for different models: All features with LOPO, binary classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest	MLP
ROC AUC	0.6556	0.6001	0.6773	0.5316	0.6209	0.5313
PR AUC	0.8806	0.8287	0.8998	0.8214	0.8630	0.8425

From the results, LightGBM shows the highest ROC AUC (0.6773) and PR AUC (0.8998), indicating better performance in distinguishing between the two classes compared to other models. Gradient Boosting and Random Forest also perform reasonably well, with ROC AUC values of 0.6556 and 0.6209, respectively. SVM and MLP perform the worst in terms of ROC AUC, indicating they might struggle more with the variability in the patient data.

2) Binary Classification with 5-folds: As mentioned above, we also use 5-fold cross-validation to investigate the performance of the models. In 5-fold cross-validation, the dataset is divided into 5 subsets, and the model is trained and tested k times, each time using a different subset as the validation set and the remaining subsets for training, providing a thorough assessment of the model's performance. This method helps to mitigate overfitting and ensures that the model is not overly dependent on any particular subset of the data.

Table IX provides the performance metrics for the same machine learning models but evaluated using 5-fold cross-validation. In this method, the dataset is split into five equal parts, and the model is trained on four parts and tested on the remaining one. This process is repeated five times, with each part used exactly once as the test set.

TABLE IX: Performance for different models: All features with 5-fold, binary classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest	MLP
ROC AUC	0.6292	0.5462	0.6892	0.5571	0.6257	0.5182
PR AUC	0.8529	0.8255	0.8833	0.8184	0.8597	0.8318

In this evaluation, LightGBM again outperforms other models with a ROC AUC of 0.6892 and a PR AUC of 0.8833. Gradient Boosting and Random Forest show comparable ROC AUC values of 0.6292 and 0.6257, respectively. Catboost and SVM exhibit lower performance, while MLP remains the lowest-performing model based on ROC AUC.

3) Multi-class Classification with LOPO: In addition to the binary classification, we also investigate the performance of models' prediction using multi-level severity following DAS21. Table X details the performance of the models on multi-class classification tasks using the LOPO cross-validation. The approach is even more challenging in a multi-class setting as the model must correctly classify multiple classes for each patient left out during testing.

TABLE X: Performance for different models: All features with LOPO, multi-class classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest
Macro ROC AUC One-vs-Rest	0.4466	0.3279	0.5678	0.4208	0.3092
Macro ROC AUC One-vs-One	0.4197	0.336	0.4781	0.4237	0.2767
Macro Precision	0.1719	0.1346	0.1336	0.1311	0.1317
Recall	0.1776	0.1384	0.1493	0.1631	0.1667
F1-score	0.1698	0.1307	0.1408	0.1454	0.1472

LightGBM exhibits the best performance for multi-class classification with LOPO, achieving a Macro ROC AUC of 0.5678 in the One-vs-Rest approach and 0.4781 in the One-vs-One approach. However, all models show relatively low performance across all metrics, reflecting the difficulty of the multi-class classification task under LOPO validation.

4) Multi-class Classification with 5-folds: Similar to the LOPO for multi-class classification, we also employ 5-fold for health issue investigation. Table XI shows the performance metrics for multi-class classification using 5-fold cross-validation. This method helps mitigate the variance seen in LOPO by averaging the performance over multiple splits.

TABLE XI: Performance for different models: All features with 5-fold, multi-class classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest
Macro ROC AUC One-vs-Rest	0.4804	0.4103	0.5812	0.4412	0.4663
Macro ROC AUC One-vs-One	0.4492	0.4132	0.5057	0.4539	0.4207
Macro Precision	0.1783	0.1434	0.1554	0.1357	0.1223
Recall	0.1746	0.1539	0.1652	0.1628	0.1667
F1-score	0.1736	0.1474	0.1578	0.1465	0.1411

Table XI shows that LightGBM continues to show the highest performance with a Macro ROC AUC of 0.5812

(One-vs-Rest) and 0.5057 (One-vs-One). Gradient Boosting and SVM also perform relatively well, but all models have lower performance metrics compared to the binary classification tasks, illustrating the increased complexity of multi-class classification.

C. Multimodal Sensor Features

1) Binary Classification with LOPO: The performance metrics for different machine learning models using the LOPO approach are summarized in Table XII. The LightGBM model achieved the highest ROC AUC score of 0.698, suggesting it performed relatively better compared to using all features as illustrated in Table VIII. Gradient Boosting followed with an ROC AUC of 0.6265, indicating moderate discriminative ability. In terms of PR AUC, which measures the trade-off between precision and recall, LightGBM again stands out with a score of 0.9091, demonstrating its robustness in handling imbalanced classes. Other models including Catboost, SVM, and Random Forest showed lower ROC AUC and PR AUC scores.

TABLE XII: Performance for different models: Multimodal sensor features with LOPO, binary classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest	MLP
ROC AUC	0.6265	0.4753	0.698	0.5124	0.556	0.5034
PR AUC	0.8379	0.7556	0.9091	0.8113	0.8209	0.7855

2) Binary Classification with 5-folds: The performance metrics for the 5-fold cross-validation approach are detailed in Table XIII. Here, LightGBM also performed well, achieving an ROC AUC of 0.6601 and a PR AUC of 0.8839, highlighting its consistent performance across different validation techniques. Gradient Boosting followed with an ROC AUC of 0.6137 and a PR AUC of 0.8424, reinforcing its reliability as a robust model for this classification task. The Catboost model showed improved performance in the 5-fold scenario (ROC AUC of 0.5145) compared to LOPO, indicating that it might be better suited for general datasets rather than patient-specific variations. SVM and Random Forest had similar ROC AUC scores, around 0.5389 and 0.5607 respectively, but they showed adequate precision-recall trade-offs with PR AUC scores above 0.82.

TABLE XIII: Performance for different models: Multimodal sensor features with 5-fold, binary classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest	MLP
ROC AUC	0.6137	0.5145	0.6601	0.5389	0.5607	0.5216
PR AUC	0.8424	0.7914	0.8839	0.8207	0.8261	0.7983

D. Top-10 Important Features

Although, we have features extracted from wearable devices and personal information, utilizing the top 10 important features for classification is a strategic approach aimed at enhancing model efficiency and interpretability. Utilizing the top 10 important features allows us to significantly reduce the

time and energy required for data collection and processing, thereby saving valuable resources and expediting the overall analysis workflow.

1) Binary Classification with LOPO: As shown in Table XIV, the models assessed include Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP. The results indicate that LightGBM achieved the highest ROC AUC score of 0.7041, followed by Gradient Boosting with a score of 0.6699. Catboost, SVM, Random Forest, and MLP showed moderate performance with ROC AUC scores of 0.5788, 0.578, 0.6232, and 0.5454, respectively. In addition, in terms of Precision-Recall AUC, LightGBM also led with a score of 0.9087, highlighting its superior ability to handle class imbalances and correctly identify positive instances in this binary classification task.

TABLE XIV: Performance for different models: Top 10 features with LOPO, binary classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest	MLP
ROC AUC	0.6699	0.5788	0.7041	0.578	0.6232	0.5454
PR AUC	0.8689	0.8213	0.9087	0.8591	0.8714	0.8413

2) Binary Classification with 5-folds: LightGBM consistently performed well, achieving an ROC AUC of 0.7168 and a PR AUC of 0.8852, underscoring its robustness and effectiveness across different cross-validation techniques. Gradient Boosting and Catboost also performed competitively with ROC AUC scores of 0.6594 and 0.6173, respectively, and PR AUC scores of 0.8723 and 0.8512.

TABLE XV: Performance for different models: Top 10 features with 5-fold, binary classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest
ROC AUC	0.6594	0.6173	0.7168	0.5692	0.6402
PR AUC	0.8723	0.8512	0.8852	0.841	0.8754

3) Multi-class Classification with LOPO: We also conducted multi-class classification training using the LOPO method. As shown in Table XVI, the performance metrics indicate a notable variation among the machine learning models. LightGBM emerged as the top performer with a Macro ROC AUC score of 0.633 (One-vs-Rest) and 0.5244 (One-vs-One), demonstrating its capability to handle multiple classes effectively. Gradient Boosting and Catboost showed moderate performance with Macro ROC AUC scores around 0.4946 and 0.4463, respectively. However, the overall macro precision, recall, and F1-score for all models were relatively low, highlighting the complexity and challenge of multi-class classification tasks using LOPO.

4) Multi-class Classification with 5-folds: Finally, we conducted multi-class classification using the same models with 5-fold cross-validation. Table XVII shows that LightGBM again led with a Macro ROC AUC score of 0.6412 (One-vs-Rest) and 0.5585 (One-vs-One), reinforcing its consistent performance across different evaluation methods. Gradient Boosting

TABLE XVI: Performance for different models: Top 10 features with LOPO, multi-class classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest
Macro ROC AUC **One-vs-Rest**	0.4946	0.4463	0.633	0.4466	0.3352
Macro ROC AUC **One-vs-One**	0.4933	0.4084	0.5244	0.4344	0.3007
Macro Precision	0.1935	0.1558	0.1636	0.1335	0.1317
Recall	0.2182	0.1737	0.1742	0.1552	0.1667
F1-score	0.1947	0.159	0.1679	0.1429	0.1472

and Catboost also showed improved performance with Macro ROC AUC scores of 0.5418 and 0.5315, respectively.

TABLE XVII: Performance for different models: Top 10 features with 5-fold, multi-class classification.

Model	Gradient Boosting	Catboost	LightGBM	SVM	Random Forest
Macro ROC AUC **One-vs-Rest**	0.5418	0.5315	0.6412	0.5022	0.3857
Macro ROC AUC **One-vs-One**	0.5507	0.4474	0.5585	0.4755	0.3615
Macro Precision	0.2314	0.1401	0.1778	0.1525	0.2896
Recall	0.2347	0.1476	0.2031	0.1705	0.1711
F1-score	0.2224	0.1416	0.1885	0.1596	0.1502

When employing the top 10 features, the binary classification performance under the LOPO scheme shows a slightly better performance in ROC AUC and PR AUC metrics across most models compared to using all features. For example, Gradient Boosting's ROC AUC increased from 0.6556 to 0.6699, while LightGBM's PR AUC slightly increased from 0.8998 to 0.9087. Similarly, in multi-class classification, the LOPO results show that models trained with the top 10 features generally have higher Macro ROC AUC and precision scores compared to those trained with all features. By focusing on the top ten important features, we can not only enhance model performance but also significantly reduce the time and energy required for data collection and processing, making the analysis more efficient and cost-effective.

E. Health Issue Explanation

While in the previous sections, we focused on the performance of various machine learning models for health issue prediction, it's crucial to understand the underlying factors influencing these predictions. This is where Explainable Artificial Intelligence (XAI) techniques come into play. XAI methods enable us to gain insights into the decision-making processes of machine learning models, offering valuable explanations for their predictions. As discussed in Section IV-A to Section IV-D, LightGBM outperforms other machine learning models. Therefore, we employed XAI techniques to interpret the top-performing LightGBM model for stress detection.

A plot of the SHAP values is illustrated in Fig. 5a, in which the features are listed on the left-hand side of the plot, with the most important features at the top. Higher SHAP values indicate a greater impact on the model output. In addition, the blue color represents the normal class and the red color is the stress class. As we can see, BMI index is the most important

feature, followed by age, gender, and heart rate. As we can see, the contribution of each feature in each class is mostly equal.

To understand the distribution of each feature in each class, we plot the SHAP values of each class in Fig. 5b and Fig. 5c. Similar to Fig. 5a, the images show scatter plots of the effects of factors on model output for each class. The x-axis represents the feature value, and the y-axis represents the SHAP value. As observed a low BMI index is associated with a lower likelihood of being classified as stressed by the model. Similar to the BMI index, Fig. 5c shows high age and low heart rate are indicative of a lower likelihood of being stressed according to the model. In addition, females seem to be more stressed than males.

V. CONCLUSION

In this study, we significantly advance mental health assessment through a three-pronged approach. First, we establish a large, novel wearable device data repository containing physiological signals and corresponding DAS-21 scores, providing a valuable resource for researchers. To our best knowledge, this is the largest and the most generalized dataset ever collected for both LDF and FS studies. Secondly, we also evaluate various machine learning models for predicting DAS levels, prioritizing interpretable models to enhance understanding of the relationship between wearable data and mental health. Finally, we employed explainable AI techniques to ensure transparency by identifying features that most influence predictions, providing insights that can help clinicians tailor treatment plans and improve patient outcomes.

Our findings show that the LightGBM model consistently outperforms others in both binary and multi-class stress level predictions, balancing accuracy and interpretability, making it suitable for practical applications. Key physiological features like heart rate, BMI, and weight significantly influence stress predictions. Using the top 10 important features, LightGBM achieved an ROC AUC of 0.7168 and a PR AUC of 0.8852.

REFERENCES

[1] Y. Wu, L. Wang, M. Tao, H. Cao, H. Yuan, M. Ye, X. Chen, K. Wang, and C. Zhu, "Changing trends in the global burden of mental disorders from 1990 to 2019 and predicted levels in 25 years," *Epidemiology and Psychiatric Sciences*, vol. 32, p. e63, 2023.

[2] S. Krokstad, D. A. Weiss, M. A. Krokstad, V. Rangul, K. Kvaløy, J. M. Ingul, O. Bjerkeset, J. Twenge, and E. R. Sund, "Divergent decennial trends in mental health according to age reveal poorer mental health for young people: repeated cross-sectional population-based surveys from the hunt study, norway," *BMJ open*, vol. 12, no. 5, p. e057654, 2022.

[3] J. Dykxhoorn, D. Osborn, K. Walters, J. Kirkbride, S. Gnani, and A. Lazzarino, "Temporal patterns in the recorded annual incidence of common mental disorders over two decades in the united kingdom: A primary care cohort study," *Psychological Medicine*, vol. 54, no. 4, pp. 663–674, 2024.

[4] K. W. Lange, "Coronavirus disease 2019 (covid-19) and global mental health," *Global health journal*, vol. 5, no. 1, pp. 31–36, 2021.

[5] L. Kola, B. A. Kohrt, C. Hanlon, J. A. Naslund, S. Sikander, M. Balaji, C. Benjet, E. Y. L. Cheung, J. Eaton, P. Gonsalves *et al.*, "Covid-19 mental health impact and responses in low-income and middle-income countries: reimagining global mental health," *The Lancet Psychiatry*, vol. 8, no. 6, pp. 535–550, 2021.
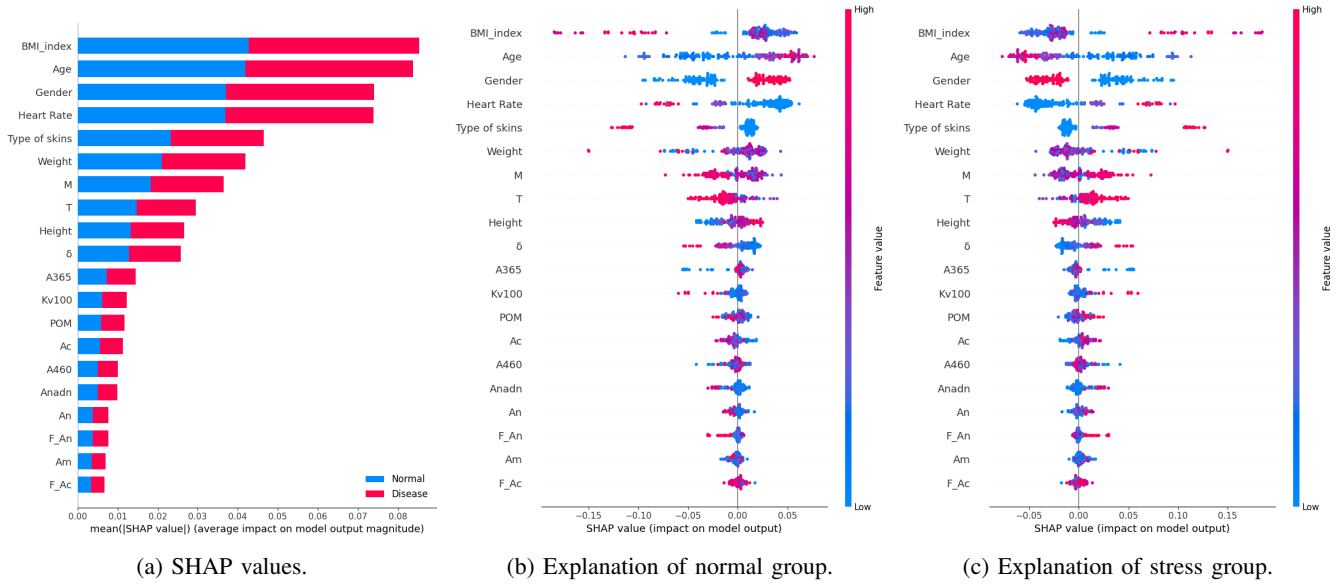
(a) SHAP values.   (b) Explanation of normal group.   (c) Explanation of stress group.

Fig. 5: Explanation of the reasoning behind individual class predictions using SHAP values.

[6] E. K.-W. Carl Baker, "Mental health statistics: prevalence, services and funding in england," *commonslibrary.parliament.uk*, vol. Number CBP-06988, pp. 1–44, 2024.

[7] G. P. Chrousos, "Stress and disorders of the stress system," *Nature reviews endocrinology*, vol. 5, no. 7, pp. 374–381, 2009.

[8] A. Steptoe and M. Kivimäki, "Stress and cardiovascular disease," *Nature Reviews Cardiology*, vol. 9, no. 6, pp. 360–370, 2012.

[9] F. Calabrese, R. Molteni, G. Racagni, and M. A. Riva, "Neuronal plasticity: a link between stress and mood disorders," *Psychoneuroendocrinology*, vol. 34, pp. S208–S216, 2009.

[10] W. Umer, Y. Yu, and M. F. Antwi Afari, "Quantifying the effect of mental stress on physical stress for construction tasks," *Journal of Construction Engineering and Management*, vol. 148, no. 3, p. 04021204, 2022.

[11] A. Pedersen, R. Zachariae, and D. H. Bovbjerg, "Influence of psychological stress on upper respiratory infection—a meta-analysis of prospective studies," *Psychosomatic medicine*, vol. 72, no. 8, pp. 823–832, 2010.

[12] M. L. Harris, C. Oldmeadow, A. Hure, J. Luu, D. Loxton, and J. Attia, "Stress increases the risk of type 2 diabetes onset in women: A 12-year longitudinal study using causal modelling," *PloS one*, vol. 12, no. 2, p. e0172126, 2017.

[13] S. H. Lovibond and P. F. Lovibond, "Depression anxiety stress scales," *Psychological Assessment*, 1995.

[14] A. T. Beck, R. A. Steer, and G. K. Brown, "Beck depression inventory," *San Antonio, TX*, 1987.

[15] A. T. Beck, N. Epstein, G. Brown, and R. Steer, "Beck anxiety inventory," *Journal of consulting and clinical psychology*, 1993.

[16] P. Miranda, C. D. Cox, M. Alexander, S. Danev, and J. R. Lakey, "Overview of current diagnostic, prognostic, and therapeutic use of eeg and eeg-based markers of cognition, mental, and brain health," *Integrative Molecular Medicine*, vol. 6, pp. 1–9, 2019.

[17] L. A. Whitten, "Functional magnetic resonance imaging (fmri): An invaluable tool in translational neuroscience," 2012.

[18] L. Wang, Y. Hu, N. Jiang, and A. K. Yetisen, "Biosensors for psychiatric biomarkers in mental health monitoring," *Biosensors and Bioelectronics*, p. 116242, 2024.

[19] R. P. Monteiro, G. L. d. H. Coelho, P. H. Hanel, V. V. Gouveia, and R. Vilar, "The 12-item mini-dass: A concise and efficient measure of depression, anxiety, and stress," *Applied Research in Quality of Life*, vol. 18, no. 6, pp. 2955–2979, 2023.

[20] J. L. McGaugh, *Emotions and bodily responses: A psychophysiological approach*. Academic Press, 2013.

[21] E. A. Zherebtsov, E. V. Zharkikh, I. Kozlov, A. I. Zherebtsova, Y. I. Loktionova, N. B. Chichkov, I. E. Rafailov, V. V. Sidorov, S. G. Sokolovski, A. V. Dunaev *et al.*, "Novel wearable vcsel-based sensors

for multipoint measurements of blood perfusion," in *Dynamics and Fluctuations in Biomedical Photonics XVI*, vol. 10877. SPIE, 2019, pp. 38–41.

[22] D. A. Low, H. Jones, N. T. Cable, L. M. Alexander, and W. L. Kenney, "Historical reviews of the assessment of human cardiovascular function: interrogation and understanding of the control of skin blood flow," *European Journal of Applied Physiology*, vol. 120, pp. 1–16, 2020.

[23] L. Kralj and H. Lenasi, "Wavelet analysis of laser doppler microcirculatory signals: Current applications and limitations," *Frontiers in Physiology*, vol. 13, p. 1076445, 2023.

[24] Y. I. Loktionova, E. A. Zherebtsov, E. V. Zharkikh, I. O. Kozlov, A. I. Zherebtsova, V. V. Sidorov, S. G. Sokolovski, I. E. Rafailov, A. V. Dunaev, and E. U. Rafailov, "Studies of age-related changes in blood perfusion coherence using wearable blood perfusion sensor system," in *European Conference on Biomedical Optics*. Optica Publishing Group, 2019, p. 11075_6.

[25] E. Zharkikh, V. Dremin, E. Zherebtsov, A. Dunaev, and I. Meglinski, "Biophotonics methods for functional monitoring of complications of diabetes mellitus," *Journal of biophotonics*, vol. 13, no. 10, p. e202000203, 2020.

[26] E. A. Zherebtsov, E. V. Zharkikh, I. O. Kozlov, Y. I. Loktionova, A. I. Zherebtsova, I. E. Rafailov, S. G. Sokolovski, V. V. Sidorov, A. V. Dunaev, and E. U. Rafailov, "Wearable sensor system for multipoint measurements of blood perfusion: pilot studies in patients with diabetes mellitus," in *European Conference on Biomedical Optics*. Optica Publishing Group, 2019, p. 11079_62.

[27] E. Zharkikh, Y. Loktionova, V. Sidorov, A. Krupatkin, G. Masalygina, and A. Dunaev, "Control of blood microcirculation parameters in therapy with alpha-lipoic acid in patients with diabetes mellitus," *Human Physiology*, vol. 48, pp. 456–464, 09 2022.

[28] E. V. Zharkikh, Y. I. Loktionova, A. A. Fedorovich, A. Y. Gorshkov, and A. V. Dunaev, "Assessment of blood microcirculation changes after covid-19 using wearable laser doppler flowmetry," *Diagnostics*, vol. 13, no. 5, p. 920, 2023.

[29] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, no. 1, pp. 381–386, 2020.

[30] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.

[31] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[32] V. Patel, A. J. Flisher, S. Hetrick, and P. McGorry, "Mental health of

young people: a global public-health challenge," *The lancet*, vol. 369, no. 9569, pp. 1302–1313, 2007.

[33] G. J. Meyer, S. E. Finn, L. D. Eyde, G. G. Kay, K. L. Moreland, R. R. Dies, E. J. Eisman, T. W. Kubiszyn, and G. M. Reed, "Psychological testing and psychological assessment: A review of evidence and issues." *American psychologist*, vol. 56, no. 2, p. 128, 2001.

[34] C. Bowman, G. Branjerdporn, K. Turner, M. Kamara, N. Tyagi, N. J. D. Reyes, and N. J. Stapelberg, "The impact of viral epidemics and pandemics on acute mental health service use: an integrative review," *Health psychology review*, vol. 15, no. 1, pp. 1–33, 2021.

[35] L. Peters, A. Peters, E. Andreopoulos, N. Pollock, R. L. Pande, and H. Mochari-Greenberger, "Comparison of dass-21, phq-8, and gad-7 in a virtual behavioral health care setting," *Heliyon*, vol. 7, no. 3, 2021.

[36] W. H. Organization *et al.*, "The impact of the covid-19 pandemic on noncommunicable disease resources and services: results of a rapid assessment," 2020.

[37] R. D. Restrepo, M. T. Alvarez, L. D. Wittnebel, H. Sorenson, R. Wettstein, D. L. Vines, J. Sikkema-Ortiz, D. D. Gardner, and R. L. Wilkins, "Medication adherence issues in patients treated for copd," *International journal of chronic obstructive pulmonary disease*, vol. 3, no. 3, pp. 371–384, 2008.

[38] Z. Yonel, P. Sharma, A. Yahyouche, Z. Jalal, T. Dietrich, and I. L. Chapple, "Patients' attendance patterns to different healthcare settings and perceptions of stakeholders regarding screening for chronic, non-communicable diseases in high street dental practices and community pharmacy: a cross-sectional study," *BMJ open*, vol. 8, no. 11, p. e024503, 2018.

[39] P. Mularczyk-Tomczewska, A. Żarnowski, M. Gujski, J. Sytnik-Czetwertyński, I. Pańkowski, R. Smoliński, and M. Jankowski, "Preventive health screening during the covid-19 pandemic: a cross-sectional survey among 102,928 internet users in poland," *Journal of Clinical Medicine*, vol. 11, no. 12, p. 3423, 2022.

[40] A. Bagno and R. Martini, "Wavelet analysis of the laser doppler signal to assess skin perfusion," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 7374–7377.

[41] A. Goltsov, A. V. Anisimova, M. Zakharkina, A. I. Krupatkin, V. V. Sidorov, S. G. Sokolovski, and E. Rafailov, "Bifurcation in blood oscillatory rhythms for patients with ischemic stroke: A small scale clinical trial using laser doppler flowmetry and computational modeling of vasomotion," *Frontiers in physiology*, vol. 8, p. 251012, 2017.

[42] A. I. Zherebtsova, V. V. Dremin, I. Makovik, E. A. Zherebtsov, A. V. Dunaev, A. N. Goltsov, S. Sokolovski, and E. U. Rafailov, "Multimodal optical diagnostics of the microhaemodynamics in upper and lower limbs," *Frontiers in Physiology*, vol. 10, 2019.

[43] E. Zherebtsov, E. Zharkikh, Y. Loktionova, A. Zherebtsova, V. Sidorov, E. Rafailov, and A. Dunaev, "Wireless dynamic light scattering sensors detect microvascular changes associated with ageing and diabetes," *IEEE transactions on bio-medical engineering*, vol. 70, pp. 3073–3081, 05 2023.

[44] V. Ralevic, A. Belai, and G. Burnstock, "Effects of streptozotocin-diabetes on sympathetic nerve, endothelial and smooth muscle function in the rat mesenteric arterial bed," *European journal of pharmacology*, vol. 286, no. 2, pp. 193–199, 1995.

[45] A. Humeau, A. Koïtka, P. Abraham, J.-L. Saumet, and J.-P. L'Huillier, "Spectral components of laser doppler flowmetry signals recorded in healthy and type 1 diabetic subjects at rest and during a local and progressive cutaneous pressure application: scalogram analyses," *Physics in Medicine & Biology*, vol. 49, no. 17, p. 3957, 2004.

[46] Y.-K. Jan, S. Shen, R. D. Foreman, and W. J. Ennis, "Skin blood flow response to locally applied mechanical and thermal stresses in the diabetic foot," *Microvascular research*, vol. 89, pp. 40–46, 2013.

[47] Y. I. Loktionova, E. V. Zharkikh, I. O. Kozlov, E. A. Zherebtsov, S. A. Bryanskaya, A. I. Zherebtsova, V. V. Sidorov, S. G. Sokolovski, A. V. Dunaev, and E. U. Rafailov, "Pilot studies of age-related changes in blood perfusion in two different types of skin," in *Saratov Fall Meeting 2018: Optical and Nano-Technologies for Biology and Medicine*, vol. 11065. SPIE, 2019, pp. 184–188.

[48] M. Saha, V. Dremin, I. Rafailov, A. Dunaev, S. Sokolovski, and E. Rafailov, "Wearable laser doppler flowmetry sensor: a feasibility study with smoker and non-smoker volunteers," *Biosensors*, vol. 10, no. 12, p. 201, 2020.

[49] E. Zharkikh, Y. Loktionova, A. Zherebtsova, M. Tsyganova, E. Zherebtsov, and A. Tiselko, "Skin blood perfusion and fluorescence parameters in pregnant women with type 1 diabetes mellitus," 10 2021, pp. 238–240.

APPENDIX A
ACKNOWLEDGEMENT

APPENDIX B
CLINICAL MOTIVATION

Over the past two decades, the incidences of Common Mental Disorders (CMDs) in global has shown significant fluctuations. According to a study published in "Psychological Medicine" [1], the rate of CMDs, particularly anxiety and depression, has increased substantially, reflecting improved awareness and diagnosis in healthcare settings [1]. However, this increase is not uniform across all age groups, with higher rates observed among younger individuals due to changing social pressures and lifestyle factors [2], [32]. Additionally, external factors such as economic conditions and public health crises have influenced mental health trends. These findings underscore the need for adaptable and accessible mental health services within the healthcare system [3]. Mental health has gained significant attention, particularly after the COVID-19 pandemic, which exacerbated mental health issues [4], [5].

In the United Kingdom, over 25% of individuals experience a mental health disorder annually, with 1 in 6 adults facing anxiety or depression weekly [4], Stress leading to overeating (46%), increased alcohol consumption (29%), and elevated smoking rates (16%) [5]. CMDs also harm various body systems, including the cardiovascular system (e.g., raising blood pressure and heart risks), nervous system (e.g., impairing learning and mood), muscles (e.g., causing tension and fatigue), breathing (e.g., resulting in shallow breathing), and metabolism (e.g., leading to weight changes and diabetes risk). Ultimately, stress affects both mental and physical well-being extensively [6].

Stress can have a detrimental impact on various body systems [7]. Prolonged stress can elevate blood pressure and heart rate, increasing the risk of cardiovascular diseases [8]. It also affects the nervous system, leading to cognitive decline, mood disorders, and an increased risk of mental disorders [9]. Muscular tension, soreness, and fatigue can result from stress, impairing daily activities [10]. Changes in breathing patterns due to stress can lead to respiratory issues [11]. Additionally, stress disrupts metabolism, potentially causing weight changes and increasing the risk of diabetes [12]. In conclusion, stress negatively affects both mental and physical health, impacting systems such as cardiovascular, nervous, muscular, respiratory, and metabolic.

Mental health assessment encompasses various methods to ensure a comprehensive and accurate understanding. Standardized tests like DAS-21 [13], the Beck Depression Inventory (BDI) [14], and the Beck Anxiety Inventory (BAI) [15] are commonly used to measure levels of depression, anxiety, and stress. Clinical interviews can be conducted in structured, semi-structured, or unstructured formats, helping psychologists gather detailed information about the patient's condition through specific questions and conversations. Behavioral observation is an important method, helping identify psychological manifestations through the patient's behaviors and reactions in various situations.

Other psychological tests include evaluations of memory, logical thinking, and concentration abilities, aiming to identify factors affecting mental health. Self-report assessments involve patients evaluating their mental state through questionnaires on quality of life, stress, and anxiety levels. Additionally, information from family and friends provides supplementary insights into the patient's mental health condition, contributing to a more accurate assessment [33].

Biological assessments also play a crucial role, including tests for neurotransmitter levels in the brain such as serotonin and dopamine, electroencephalograms (EEGs) to monitor brain electrical activity [16], and functional magnetic resonance imaging (fMRI) to observe brain activity during psychological tasks [17]. The methods using biosensors for monitoring psychiatric biomarkers (cortisol, dopamine, and serotonin) are introduced as potential tools for diagnosing and managing psychiatric disorders. It is possible to collect psychiatric biomarkers like cortisol, dopamine, and serotonin through blood, saliva, urine, and sweat samples. Advantages include high sensitivity, selectivity, and real-time monitoring capabilities. However, limitations involve accuracy being affected by environmental factors, high costs, and challenges in integrating data from multiple sensors for comprehensive diagnosis. Continuous development and improvement are required to enhance their effectiveness in mental healthcare applications [18].

Screening for depression and anxiety has increased, especially with remote methods during COVID-19, allowing quick identification before hospital referrals. Digital tools like apps and online platforms provide cost-effective mental health support but face challenges like inconsistent engagement, variable quality, and data security issues. Despite these, they show promise and need further development and standardization [34].

Consequently, the Depression Anxiety Stress Scales (DASS) have emerged as one of the most widely used tools to assess negative emotional states. As of April 2023, the article by [13] had garnered 13,130 citations on Google Scholar. A useful tool

---

[4]https://www.mind.org.uk/news-campaigns/news/mind-urges-the-nation-speak-to-us-during-mental-health-awareness-week/
[5]https://www.myndup.com/blog/mental-health-statistics-2023

for this purpose is the DAS-21 questionnaire, a short version of the 42-item DASS. The DAS-21 consists of 21 items divided into three subscales: Depression, Anxiety, and stress. The Depression subscale assesses loss of motivation and self-esteem, the Anxiety subscale measures persistent anxiety symptoms, and the stress subscale addresses symptoms of irritability and tension. The DAS-21 has been validated in both clinical and community settings through standard structured clinical interviews and has demonstrated excellent internal consistency, with Cronbach's α values of 0.94 for Depression, 0.87 for Anxiety, and 0.91 for stress. The severity levels and cutoff points of the DAS-21 help classify and promptly support patients [19], [35].

In addition, intense emotions can affect the hands through the interaction between the nervous system and mood. When experiencing emotions such as anxiety or anger, the body can react by altering blood flow intensity and muscular electrical activity in the hands, leading to muscle tension or relaxation [20]. This can impact the ability to use the hands in daily activities, such as writing, drawing, or performing tasks. Additionally, the hands can become sensitive and express emotional expressions through gestures and usage.

Although there are many articles on blood circulation in people with this condition, stress was measured using the previously described 2-channel multi-mode wearable portable wireless diagnostic device (LDF and FS), no studies have focused on the comparison of the variability of the parameters of blood circulation in people with and without stressful conditions. Therefore, this study demonstrated the feasibility of the wearable device in differentiating cardiovascular parameters between the stress and non-stress groups on the both middle fingers of the left and right hand.

APPENDIX C
LITERATURE REVIEW

Global health faces dual challenges from infectious diseases like COVID-19 and rising non-communicable diseases (NCDs). The World Health Organization (WHO) report highlights that the COVID-19 pandemic has caused significant disruptions in chronic disease services. Specifically, 53% of countries reported disruptions in hypertension treatment, 49% in diabetes care, 42% in cancer treatment, and 31% in cardiovascular emergency services. Additionally, over 50% of countries postponed public screening programs for breast and cervical cancer due to the reassignment of healthcare staff to COVID-19 duties and the cancellation of planned treatments [36]–[39]. The use of remote monitoring devices without intervention is crucial to aid patients and healthcare professionals in timely classification and treatment. These devices can continuously monitor vital health indicators, detect abnormalities early, and alleviate the burden on the healthcare system.

A promising new wearable technology:

Fortunately, advancements in technology offer promising solutions. Professor E. Rafailov's research group at Aston University has made significant strides in developing LDF/FS wearable devices [21]. They developed the devices using VCSEs, which demonstrated signal responses comparable to conventional tabletop monitors through volunteer-based assessments. These devices use Laser Doppler Flowmetry (LDF) and fluorescence spectroscopy (FS) for non-invasive early detection of vascular complications in diabetes and other conditions. LDF assesses tissue perfusion, oxygen saturation, and blood volume by analyzing backscatter from red blood cells using near-infrared and infrared light. FS complements LDF by detecting metabolic activity changes and advanced glycation end-products (AGEs) accumulation in diabetes, which contributes to microvascular damage and inflammation.

The benefits and applications of the wearable LDF/FS:

LDF is a non-invasive method for estimating perfusion in the microcirculation [22]. The technique was first introduced over 30 years ago and has undergone continuous development since then. The principle involves probing tissue with laser radiation and analyzing the backscatter from the partially scattered tissue radiation caused by the movement of red blood cells (whose main component is Hemoglobin - Hb). The main parameter recorded by this method is called the microcirculation or perfusion index. The microcirculatory system is the main structural and functional unit of organs, providing nutrition, adaptation, and regulation for their micro-regions.

The scientific concept behind LDF:

The method uses wavelet transformation to assess microvessel oscillatory processes over a wide frequency range. For over 15 years, adaptive wavelet analysis with complex-valued Morlet wavelets has been standard for evaluating microvascular fluctuations, previously utilizing fast Fourier transform and Butterworth filters [23]. Since LDF-gram (perfusion) is non-stationary, continuous wavelet transformation is preferred for its optimal "time-frequency" resolution, effectively tracking frequency and amplitude fluctuations in blood flow signals [24].

Specific details of analyzing the microcirculation using LDF and wavelet transformation:

Currently, several principal frequency bands are distinguished in microvascular oscillations, reflecting various regulatory mechanisms: endothelial 0.0095–0.02 Hz, neurogenic 0.02–0.06 Hz, myogenic 0.06–0.16 Hz, respiratory 0.16–0.4 Hz, and cardiac 0.4–1.6 Hz [40].

The concept behind Fluorescence Spectroscopy (FS):

The FS method uses laser probing to record the fluorescence spectra of metabolic coenzymes, measuring the normalized amplitude of NADH and FAD fluorescence intensity. This determines changes in metabolic activity in endothelial cells, which

can indicate various physiological and pathological processes. Thus, fluorescence monitoring detects cellular metabolic disorders related to different diseases [25].

Additional capabilities of Fluorescence Spectroscopy (FS):

Additionally, other structural proteins of capillary and skin membranes also exhibit fluorescence: pentosidine residues formed during collagen glycation. Pathogenic factors such as hyperglycemia and oxidative stress in diabetes lead to increased protein glycation and accumulation of advanced glycation end products (AGEs), affecting the properties of collagen with specific wavelengths of light. This can be used to study skin fluorescence related to AGE accumulation, which is associated with the accumulation of these substances.

Currently, there is a growing interest in wearable electronic diagnostic devices because daily monitoring of parameters promises a new quality of diagnosis. Recently, multimodal approaches have been actively developed, allowing clinicians to obtain in vivo values of physiological and biochemical parameters, as well as to comprehensively assess the viability of the subcutaneous microcirculatory system. One of the first developments of wearable devices for estimating subcutaneous microcirculatory tissue system (MTS) parameters is the "LAZMA PF" analyzer, produced by Aston Medical Technology Ltd., UK, under the name "FED-1B". This device integrates a multimodal approach, specifically including 2 channels for laser Doppler flowmetry (LDF) and fluorescence spectroscopy (FS), designed into a new device named MDFED-2B[6].

This technology is renowned for its non-invasive measurement capabilities in living tissues. Studies have been conducted at various sites such as the wrist, ankle, thigh, and fingertips. It has many applications, including research on metabolic and vascular complications of diabetes, automatic cerebral vascular analysis, and monitoring cerebral circulation in both healthy individuals and those with disorders. It provides continuous, non-invasive monitoring during diagnostic, treatment, and post-treatment phases. Spectral characteristic changes have been observed in conditions such as malignant tumors, surgical trauma, increased arterial pressure, and many others. It is also used to assess the functional status of the cerebral vascular system in patients with acute and chronic cerebrovascular disorders.

Monte Carlo modeling has shown a penetration depth of up to 2 mm for the LDF channel (deep vessels) and 1 mm for the FS channel . Sensors can monitor parameters such as perfusion, movement, skin temperature, and metabolic activity, providing crucial information for evaluating various physiological processes.

The technical details of the MDFED-2B wearable device and how it uses its two channels for LDF and FS measurements: The two channels combined in the design characteristics of the wearable device are used for multimodal optical diagnostics. A distinctive feature of the wearable devices under consideration is the absence of optical fibers in the design, which reduces common motion artifacts on the fibers. The wearable devices are placed on the skin for direct irradiation from a window on the underside of the device, recording the emitted (secondary) radiation from the biological tissue on the back of the device and transmitting measurement data to a PC via Bluetooth or Wi-Fi protocol. The "MDFED-2B" wearable device with 2 optical diagnostic channels uses an 850 nm VCSEL chip as the single-mode laser source with 0.8 mW power in the LDF channel, directly transmitting radiation to the skin. In the FS channel, a UV 365 nm LED is used, with a pulse power of 1.4 mW and an average power of 0.35 mW to excite endogenous NADH fluorescence at wavelengths between 460-470 nm. The amplitude of NADH fluorescence intensity (ANADH) is normalized to backscattered radiation to reduce the influence of varying blood filling in the biological tissue, which arises, among other reasons, from artifacts related to different pressures on the skin surface. The distance between the windows of the 2 channels is approximately 1 cm (the distance between the radiation source and detector). The placement of wearable devices for MTS diagnostics of the human body and wireless connection to a personal computer or smartphone is shown on symmetric regions of the limbs. Common locations for wearable devices on the body's biological tissue depend on the diagnostic task: these are usually symmetrical points on the right and left of the upper and lower limbs, areas with direct arterio-venous connections (hand or fingertip) and with predominant nutritional blood flow (forearm or lower leg), and on the forehead at the supraorbital artery regions.

The applications of the LDF/FS technology in the MDFED-2B wearable device for various medical conditions:

Studies have applied this technology to measurements in various patients. Low perfusion parameters were observed in 19 patients with acute ischemic stroke (AIS) in both the affected and unaffected hemispheres, and were lower in patients with chronic cerebrovascular disease [41]. It evaluated the severity of microcirculatory and metabolic disorders in 41 rheumatic diseases and 76 diabetes patients. Research on perfusion in patients with joint microcirculatory disorders in the hands [42] . Cardiovascular risk in diabetic and elderly patients. A study on three groups, including 37 diabetic patients, 37 elderly individuals, and 58 young individuals, comparing average perfusion using the LDF (Laser Doppler Flowmetry) method, showed that blood microcirculation index values increase with age and the progression of diabetes [43]. There was no statistically significant difference between patients and the older control group in average perfusion. However, the average energy of blood flow oscillations decreased in patients with diabetes in the endothelial, neurogenic, myogenic, and respiratory ranges . In terms of neurogenic and myogenic variability, statistically significant differences were found between the diabetic patient group and the older control group, reflecting the influence of sympathetic nervous distribution and vascular smooth muscle activity [44].

---

[6]https://amedtech.co.uk/product/mfed-2b/

In another study, diabetic patients had significantly lower endothelial, neurogenic, and myogenic low-frequency oscillation values compared to healthy controls when measured near the head of the first metatarsal bone. [45] When measuring on smooth skin, Jan and colleagues also showed reduced neurogenic and myogenic regulation in diabetes in response to heat. According to the authors, such changes in blood flow regulation are due to disruptions in the autonomic component of the peripheral nervous system, causing blood flow to be diverted to shunts.

The most common characteristic of microcirculatory disorders related to diabetes is the dysfunction of smooth muscle, endothelial cells, and perivascular nerves in the periphery, which explains the decrease in low-frequency perfusion oscillation values [43]. A study on an animal model of diabetes also showed that neurogenic distribution disorders in peripheral vessels are the primary factor contributing to microcirculatory dysfunction, leading to endothelial dysfunction and impaired smooth muscle function in the vascular system [26], [42], [46], [47]. Non-smokers had higher blood perfusion levels compared to smokers, while smokers exhibited greater variation in pulse frequency. These findings suggest that the LDF device is effective in detecting the cardiovascular impacts of smoking and could be useful for monitoring blood microcirculation and related pathologies in smokers. [48]. The device's advantages are its painlessness, quick results, no need for expensive consumables, and minimal impact on the patient.

The validation steps taken before using the LDF/FS wearable devices (MDFED-2B) for various diagnoses:

Before deploying these new wearable devices for various diagnostic tasks, their potential for multipoint perfusion measurement was investigated. These devices were used to analyze the synchronization of blood flow on the skin in analogous regions of opposite limbs, both at rest and during various functional tests (occlusion or breath-holding). Studies have shown high synchronization of blood flow rhythms in the opposite limbs of healthy volunteers. The compact and highly sensitive devices can be used even outside clinical settings. Furthermore, these wearable devices show high repeatability of measurements at rest and during physiological tests, enhancing the diagnostic value of the measurements.

Findings from an initial study using the LDF/FS wearable devices to investigate blood microcirculation:

A study using new wearable devices examined blood microcirculation across age groups. Older adults showed higher perfusion levels in the middle palm and dorsal forearm, due to thinner skin reducing laser scattering and increasing diagnostic volume. Younger individuals had higher wavelet parameters in blood perfusion oscillations. These findings can aid in developing MTS study protocols for patients with various pathologies. [24].

How the LDF/FS wearable device was used in a study on diabetes and microcirculation:

Research using wearable diagnostic devices in endocrinology assessed microcirculatory function in 19 diabetes (DM) type 2 patients and 37 healthy individuals across two age groups. Results showed different average perfusion levels between healthy volunteers of different ages and between younger healthy volunteers and DM patients. Notably, wrist and fingertip perfusion levels in healthy groups showed no significant difference. This pilot study demonstrated that wireless LDF wearable devices are convenient for point-of-care testing, recording age-specific perfusion changes and changes related to diabetes development [26].

A study using the LDF/FS wearable device to monitor the effects of alpha-lipoic acid treatment on microcirculation in diabetic patients:

Another promising use of these wearable devices in endocrinology is monitoring 10 diabetes patients therapy involving intravenous alpha-lipoic acid. Studies showed a decrease in microcirculatory and nutritional blood flow and an increase in shunt blood flow during treatment. After treatment, patients' parameters approximated control group values, particularly in the lower limbs, which are more affected by diabetic complications due to higher stress factors. These changes suggest positive effects of the therapy [27].

A study using the LDF/FS wearable device to investigate the impact of pregestational type 1 diabetes on microcirculation in pregnant women at different stages:

The study examined multimodal wearable diagnostic devices' impact on pregestational type 1 diabetes in pregnant women, showing glucose variability monitoring's role in assessing vascular function and oxidative status. Ten pregnant women (ages 32, 7-22 weeks gestation) and seven healthy women (age 32) were monitored using the "Libre Freestyle" system. Results indicated reduced microvascular activity in pregnant patients' legs and increased NADH fluorescence, suggesting tissue respiration decline [49].

A study using wearable LDF devices to analyze blood flow patterns in patients with COVID-19 during different stages of recovery:

The study demonstrated the use of peripheral blood flow oscillation analysis with wearable LDF devices to diagnose vascular disorders in a COVID-19 patient during early and progressive recovery stages. Results showed a significant increase in neurogenic oscillation amplitude in the upper limbs, potentially leading to arteriolar and venular dilation and microcirculatory blood flow shunting, adversely affecting oxygen delivery and tissue metabolism. Wavelet analysis confirmed changes in average perfusion levels due to blood flow fluctuations, influenced by disease severity and specifics [28].

The potential the wearable LDF/FS devices in medical diagnosis:

Summarizing all the given data, the presented wearable diagnostic devices with 2 optical channels for LDF and FS are a promising approach for evaluating the functional state of MTS. The multimodal approach of using LDF and FS makes it possible to simultaneously obtain physiological and metabolic information, which helps to comprehensively assess the state of the microcirculatory system. The results presented in studies demonstrated the possibility of using wearable devices to obtain objective information on MTS status under normal and pathological conditions. However, more detailed studies with larger patient cohorts and extended analysis of physiological conditions should be conducted for further clinical implementation.

One of the crucial tasks is to investigate the effects of various treatment protocols and lifestyle changes on microcirculatory and metabolic parameters using these wearable devices. Another important direction is developing machine learning algorithms for automated data analysis and interpretation, which could significantly enhance the diagnostic capabilities of wearable devices.

The development and application of wearable diagnostic devices with LDF and FS channels represent a significant advancement in medical diagnostics, offering non-invasive, real-time monitoring of the microcirculatory system and metabolic state. These devices hold great potential for improving patient care, particularly in managing chronic diseases and monitoring treatment efficacy. Our research aims to leverage this technology to build a dataset for stress detection. The volunteers in our work have diverse medical histories, including migraine, diabetes, STEMI, and hypertension. By exploring cutaneous blood microcirculation parameters using a non-invasive wearable device equipped with LDF and FS channels, we can gain valuable insights. To the best of our knowledge, our work pioneers in publishing a large LDF/FS wearable device dataset for mental health assessment.

APPENDIX D
DETAILED STUDY DESIGN AND DATASET DESCRIPTION

A. Clinical Definition and Data collection

The criteria of this study are described as follows:

- This study focuses on volunteers aged 18 and above.
- Volunteers must not have any medical conditions related to dermatological diseases on both hands and middle fingers.
- A total of 132 volunteers, aged 18 and above, of all genders, various occupations, and different ages, who are healthy and alert, were included.

Table XVIII shows the depression, anxiety and stress scale (DAS21) questionnaire responses which we gave to our volunteers.

No.	Question	0	1	2	3
1 (s)	I found it hard to wind down	0	1	2	3
2 (a)	I was aware of dryness of my mouth	0	1	2	3
3 (d)	I couldn't seem to experience any positive feeling at all	0	1	2	3
4 (a)	I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion)	0	1	2	3
5 (d)	I found it difficult to work up the initiative to do things	0	1	2	3
6 (s)	I tended to over-react to situations	0	1	2	3
7 (a)	I experienced trembling (e.g. in the hands)	0	1	2	3
8 (s)	I felt that I was using a lot of nervous energy	0	1	2	3
9 (a)	I was worried about situations in which I might panic and make a fool of myself	0	1	2	3
10 (d)	I felt that I had nothing to look forward to	0	1	2	3
11 (s)	I found myself getting agitated	0	1	2	3
12 (s)	I found it difficult to relax	0	1	2	3
13 (d)	I felt down-hearted and blue	0	1	2	3
14 (s)	I was intolerant of anything that kept me from getting on with what I was doing	0	1	2	3
15 (a)	I felt I was close to panic	0	1	2	3
16 (d)	I was unable to become enthusiastic about anything	0	1	2	3
17 (d)	I felt I wasn't worth much as a person	0	1	2	3
18 (s)	I felt that I was rather touchy	0	1	2	3
19 (a)	I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat)	0	1	2	3
20 (a)	I felt scared without any good reason	0	1	2	3
21 (d)	I felt that life was meaningless	0	1	2	3

TABLE XVIII: Depression, anxiety and stress scale (DAS21) questionnaire responses.

B. Data Analysis

Our dataset focuses on understanding the factors influencing depression, anxiety, and stress (DAS) levels. To achieve this, we have collected and integrated data from three key sources: personal information, wearable sensor readings, and the DAS-21 questionnaire. This diverse data representation allows us to create a comprehensive picture of each individual's background, psychological state, and physiological responses.

Similar to a medical record, the dataset includes essential demographic details for each participant. Our research involved a wide range of ages (18-94) with an average participant age of 40. We also ensured participant diversity by including a variety of races (Asian, White, African) and genders (55.2% male, 44.5% female), as shown in Fig. 6a and Fig. 6b, respectively.

(a) Race   (b) Gender   (c) Smoke

(d) Mental health status   (e) Sleep during experiment   (f) Blood Pressure
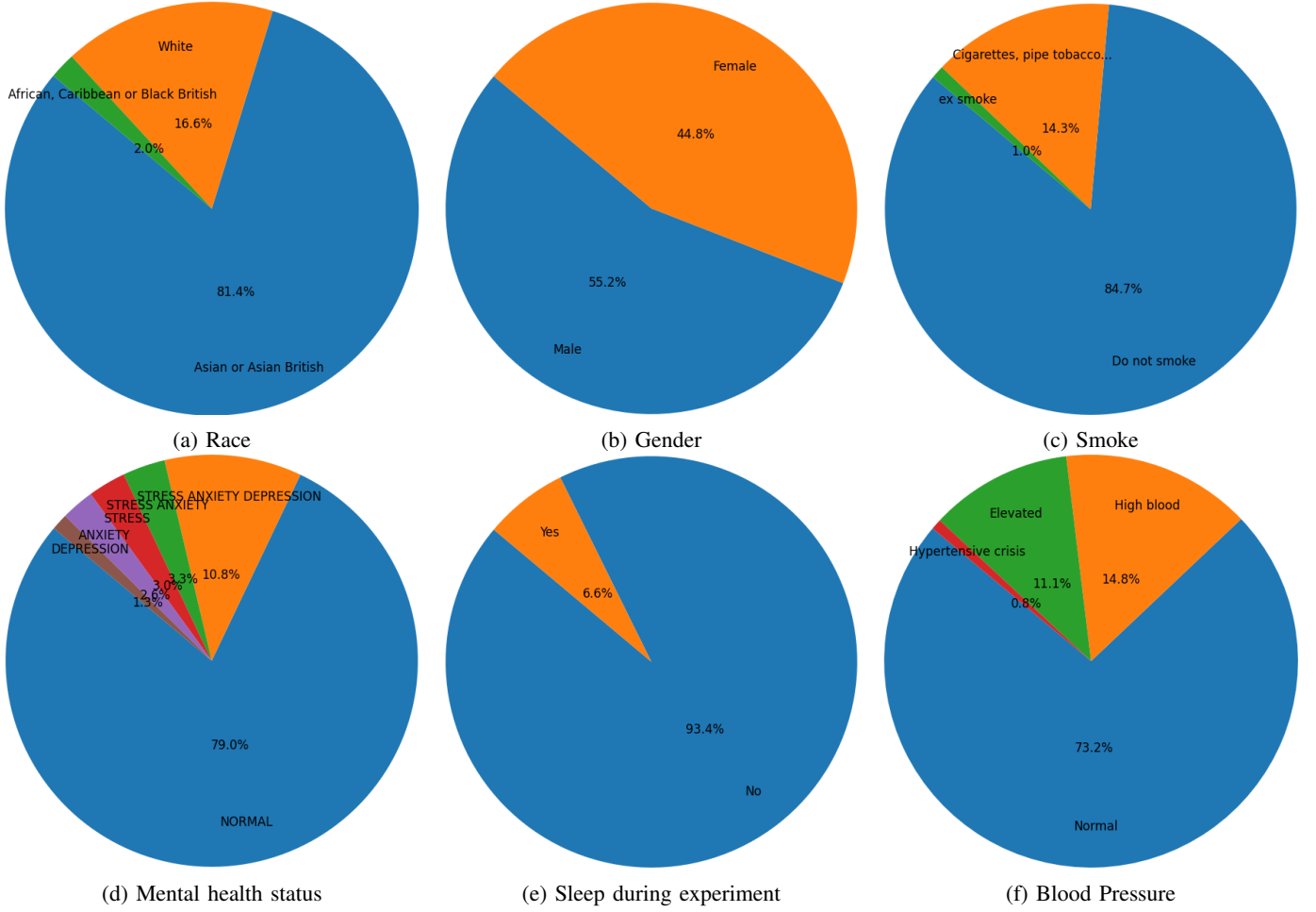
Fig. 6: Data Analysis

In addition to the demographic information, we also investigate the effect of patients' routines and medical history, which is used to further understand potential contributors to DAS. As shown in Fig 6c, most of the anticipants do not smoke cigarettes, and the rate of people used to smoke is small compared to smoked people and non-smoke people. A similar observation is seen in the health issues attribute, in which the percentage of people who have problems with health issues is significantly smaller than normal people, as illustrated in Fig. 6d.

The dataset incorporates data collected from wearable devices worn by participants. As shown in Fig. 6e, we even recorded details like sleep patterns (sleeping vs. awake) and the hand used for blood sample collection during the study. These devices continuously monitor various physiological and activity-related aspects, providing real-time health information. Examples of data collected include Body Mass Index (BMI), heart rate, and Blood Pressure. Each data point is linked to a specific time, which allows us to analyze trends and potential correlations between a participant's physiological responses and their mental state over time. The experiment duration is around 15 minutes, at the end of the experiment.

Table XIX and Fig. 7 indicate that non-wellbeing individuals exhibit higher perfusion parameters and amplitude variations compared to wellbing individuals. For the perfusion parameter M, wellbeing individuals have a mean value of 21.02 (95% CI, 4.73 - 35.59), whereas non-wellbeing individuals have a significantly higher M value of 26.49 (95% CI, 8.59 - 36.71), (p=0.016, Mann-Whitney U test). The endothelial, neural, muscle, respiratory, and cardiovascular amplitude variations all tend to be higher by more than 0.2, but the differences are not statistically significant.

As shown in Table XX and Fig. 8, non-wellbeing individuals exhibit a statistically significant higher amplitude of perfusion parameter fluctuations (δ) compared to their wellbeing counterparts, with values of 4.79 (95% CI, 2.25 - 7.65) versus 3.64
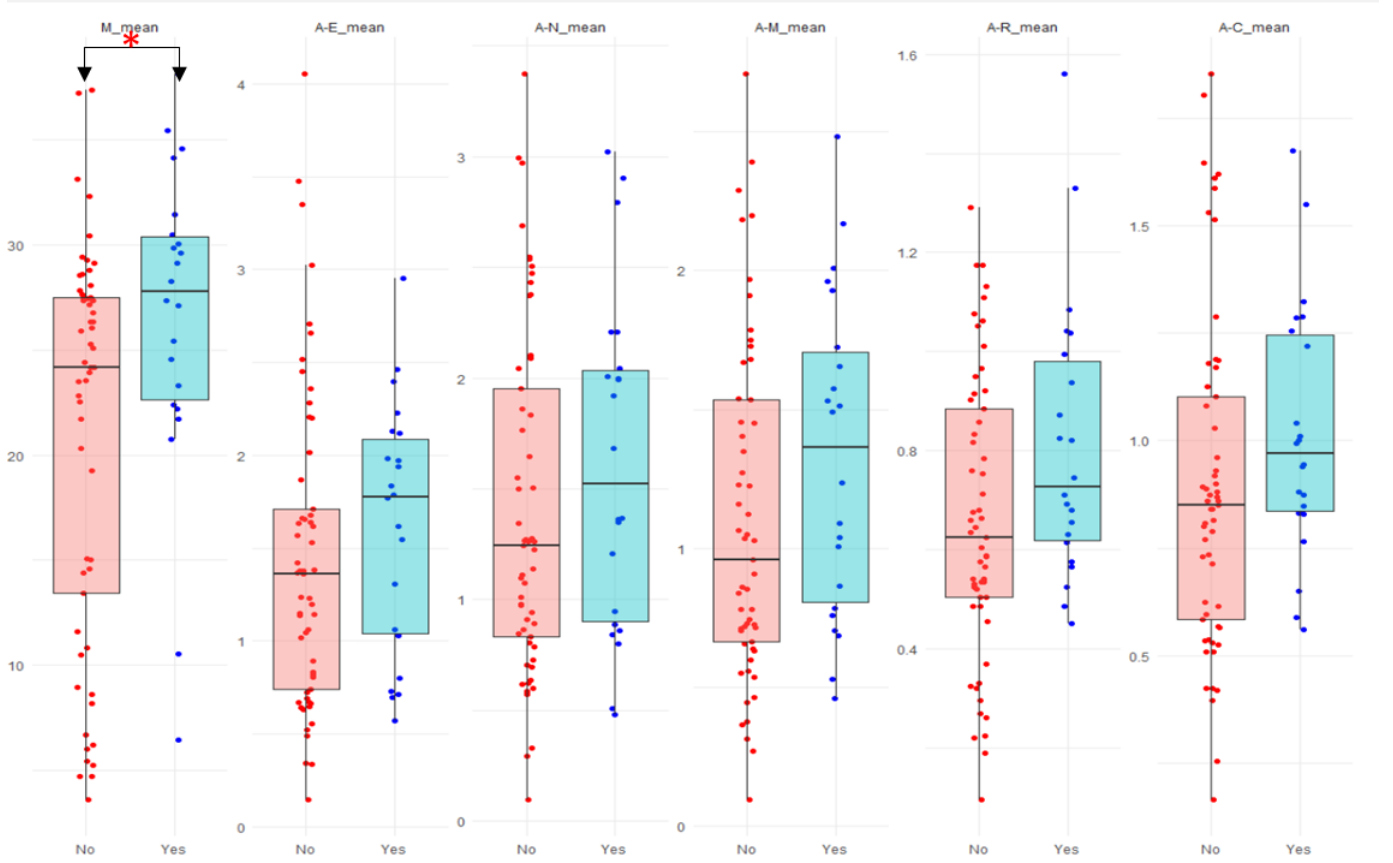
Fig. 7: Blood perfusion (M*) with standard deviation and Maximum amplitude with standard deviation of the endothelial (A-E), neurogenic (A-N), myogenic (A-M), breath (A-R) and pulse (A-C) mechanism for Wellbeing vs Non-wellbeing, *, p<0.01, Mann-Whitney U test.

Subgroup	M_mean	p-value	A-E_mean	p-value	A-N_mean	p-value	A-M_mean	p-value	A-R_mean	p-value	A-C_mean	p-value
All	22.54 (4.73 - 37.23)		1.49 (0.34 - 3.36)		1.46 (0.33 - 3)		1.17 (0.31 - 2.4)		0.7 (0.22 - 1.29)		0.93 (0.39 - 1.68)	
Wellbeing	21.02 (4.73 - 35.59)	0.016121	1.44 (0.34 - 3.42)	0.171585	1.4 (0.31 - 2.99)	0.27169	1.11 (0.29 - 2.35)	0.123042	0.66 (0.2 - 1.18)	0.0614	0.89 (0.31 - 1.74)	0.06069
Non-Wellbeing	26.49 (8.59 - 36.71)		1.62 (0.64 - 2.7)		1.61 (0.49 - 2.96)		1.33 (0.5 - 2.32)		0.81 (0.47 - 1.44)		1.02 (0.58 - 1.61)	

TABLE XIX: Blood perfusion (M*) with standard deviation and Maximum amplitude with standard deviation of the endothelial (A-E), neurogenic (A-N), myogenic (A-M), breath (A-R) and pulse (A-C) mechanism for Wellbeing vs Non-wellbeing, *, p¡0.01, Mann-Whitney U test.

(95% CI, 0.98 - 6.93), (p=0.007, Mann-Whitney U test). Additionally, they have a significantly higher temperature at the measurement site, 33.21 (30.14 - 35.82) compared to 30.65 (95% CI, 22.68 - 35.6), (p=0.01, Mann-Whitney U test). Moreover, the metabolic index (POM) at the measurement site is also significantly elevated in non-wellbeing individuals, with values of 11.42 (95% CI, 1.95 - 24.42) versus 7.7 (95% CI, 0.85 - 20.43), (p=0.01, Mann-Whitney U test).

Subgroup	Kv100_mean	δ_mean	p-value	T_mean	p-value	A365_mean	A460_mean	Anadn_mean	POM_mean	p-value
All	21.09 (6.86 - 49.55)	3.96 (1.21 - 7.41)		31.36 (22.95 - 35.79)		86.82 (4.42 - 158.6)	59.3 (12.92 - 106.52)	1.01 (0.4 - 4.54)	8.74 (0.99 - 22.15)	
Wellbeing	21.12 (6.71 - 48.75)	3.64 (0.98 - 6.93)	0.007252	30.65 (22.68 - 35.6)	0.018427	85.43 (9.5 - 130.9)	60.64 (17.2 - 106.8)	1.01 (0.41 - 4.77)	7.7 (0.85 - 20.43)	0.010656
Non-Wellbeing	21 (7.56 - 48.44)	4.79 (2.25 - 7.65)		33.21 (30.14 - 35.82)		90.43 (2.52 - 159.81)	55.81 (12.05 - 88.06)	1 (0.39 - 4.16)	11.42 (1.95 - 24.42)	

TABLE XX: The parameters with standard deviation for Wellbeing vs Non-wellbeing: Kv100, δ^*, T^*, A365, A460, Anadn, POM^*, F-E; F-N; F-M; F-R; F-C, *, $p < 0.01$, Mann-Whitney U test.
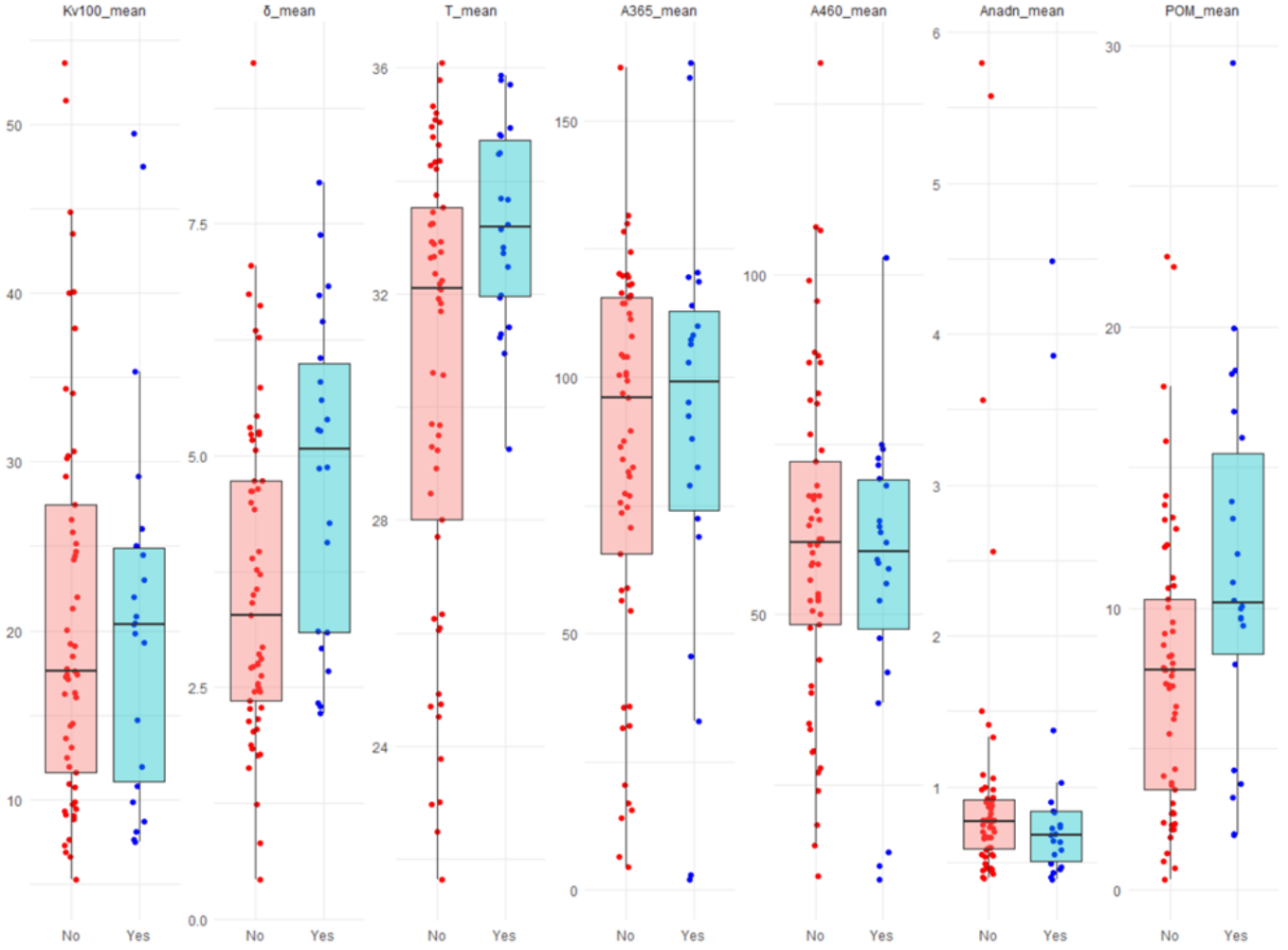
Fig. 8: The parameters with standard deviation for Wellbeing vs Non-wellbeing: Kv100, δ*, T*, A365, A460, Anadn, POM*, F-E; F-N; F-M; F-R; F-C, *, p<0.01, Mann-Whitney U test.

APPENDIX E
MACHINE LEARNING AND EXPLAINABLE ARTIFICIAL INTELLIGENCE

This section details the machine learning models employed for predicting depression, anxiety, and stress (DAS) levels, along with the Explainable Artificial Intelligence (XAI) technique used to interpret their decision-making processes.

APPENDIX F
MACHINE-LEARNING MODELS FOR DAS PREDICTION

To identify the most effective approach for predicting Depression, Anxiety, and stress (DAS) levels, we explored various machine learning algorithms. These algorithms leverage the collected wearable device data and DAS-21 questionnaire scores to estimate a patient's mental issues.

Support Vector Machine (SVM) is a well-established method that is known for its ability to effectively handle high-dimensional datasets, even with a relatively small sample size. SVMs aim to find a hyperplane in the feature space that maximizes the margin between different classes. In SVM, new data points are then classified based on which side of the hyperplane they fall on. SVMs are powerful for classification tasks with high-dimensional data, such as ours with potentially many features extracted from wearable sensors. They are effective even with limited data and offer good generalization capabilities. However, SVMs can be computationally expensive for very large datasets and may be less interpretable than other algorithms on this list.

Random Forest Classifier is an ensemble learning algorithm that combines the predictions of multiple, independently trained decision trees. Each tree is built using a random subset of features and data points, promoting diversity within the

ensemble. The final prediction is made by majority vote or averaging the individual tree predictions. The algorithms are robust to overfitting due to their inherent diversity. They can handle various data types and perform well even with missing values. This approach is particularly useful for datasets with potential noise or inconsistencies.

Gradient Boosting Classifier is an algorithm that works by iteratively building an ensemble of weak decision trees. Each tree learns from the errors of the previous one, ultimately leading to a more robust and accurate model. Gradient boosting is known for its flexibility and ability to handle various data types, making it a strong contender for our analysis.

Building upon gradient boosting, **CatBoost** specifically addresses challenges in healthcare data. It incorporates advanced techniques for handling categorical features, such as one-hot encoding or custom loss functions, which can be problematic for traditional gradient boosting. Additionally, CatBoost prioritizes model interpretability by providing feature importance scores and visualizations of decision boundaries. CatBoost excels in scenarios with a high volume of categorical features, common in healthcare data. It offers improved interpretability compared to standard gradient boosting, allowing us to understand the factors influencing model predictions

LightGBM (Light Gradient Boosting Machine) is a highly efficient implementation of the gradient boosting algorithm specifically designed for speed and performance. It utilizes techniques similar to gradient-based one-side sampling and feature importance sampling to focus on informative data points and reduce computational costs. LightGBM offers exceptional speed and accuracy, making it a compelling choice for large datasets. It is particularly efficient for memory usage, allowing for training in resource-constrained environments. LightGBM excels at handling large and complex datasets, making it suitable for our analysis where we have a high volume of data points from wearable devices.

In addition to the machine learning algorithms above, we also implemented a **Multi-layer Perceptron** (MLP) for health issue prediction. Unlike simpler models, MLPs excel at identifying complex, non-linear relationships within the data. This capability could be particularly valuable for uncovering subtle patterns between physiological signals and mental health states. The MLP neural network has two hidden layers and a final output layer with a unit. The network employs layer normalization, ReLU activation for hidden layers, and dropout to prevent overfitting. Finally, a sigmoid or a softmax activation function is applied to the output layer to transform the final values to the probability of classes.

By evaluating the performance of these diverse algorithms, we aim to identify the one that best predicts DAS levels in the context of our specific dataset and research goals.

APPENDIX G
EXPLAINABLE AI

In healthcare applications, ensuring trustworthy AI requires models to be not only accurate but also interpretable. Understanding the reasoning behind a model's predictions for DAS levels is crucial for building trust and confidence in its outputs. This empowers healthcare professionals and researchers to make informed decisions based on the predicted DAS levels and the underlying factors influencing those predictions. In this study, we leverage SHAP (Shapley Additive Explanations) to achieve interpretability and gain insights into the model's decision-making process for DAS prediction [31].

SHAP, a powerful approach for achieving interpretability, assigns an attribution value (SHAP value) to each feature for a given DAS prediction. This value represents the contribution of that specific feature (e.g., a specific physiological sensor reading or a DAS-21 questionnaire response) to the model's final prediction. High positive SHAP values indicate that the feature has a strong positive influence on the predicted DAS level (potentially indicating a higher likelihood of depression, anxiety, or stress). Conversely, low negative SHAP values signify a negative influence (potentially indicating a lower likelihood). By analyzing the SHAP values for each feature, we can gain insights into the relative importance of various factors shaping the model's predictions about a user's mental state.

This interpretability allows us to answer several key questions:

- Identification of Key Physiological and Psychological Indicators: What are the features from wearable sensor data and questionnaire scores of a patient that have the most significant influence on the model's predictions? Understanding these can pinpoint crucial physiological and psychological indicators associated with DAS levels. This knowledge can inform the development of targeted interventions and preventative measures for mental health.
- Validation of Model Fairness and Mitigation of Bias: Are the model's predictions fair across different demographics (age, gender, etc.)? Examining SHAP values across these groups helps ensure that the model is not unfairly biased towards certain populations. This is crucial in healthcare applications where fairness and unbiased decision-making are paramount.
- Enhanced Model Transparency and Trust: How does the model arrive at its predictions? By explaining the rationale behind the model's predictions through SHAP values, we can foster trust and confidence in its use among healthcare professionals and researchers. This transparency is essential for the adoption and responsible use of AI in mental health assessments.

A. Case study

Our study employs two primary approaches to train machine learning models for predicting DAS levels: binary classification and multi-class classification, as shown in Table XXI and Table XXII. Both approaches leverage data from the DAS-21 questionnaire alongside potentially other features from the collected dataset. In addition, we consider three cases to investigate the models' performances: Using all collected features as shown in Table XXI, using only features extracted from wearable devices, and using top-10 important features as illustrated in Table XXII.

TABLE XXI: Model investigation with all collected features, in which we trained models with different data split methods: 80:20, 5-fold, and leave one patient out (LOPO). In addition, "x" presents the cases that are being trained in our investigation.

Model	All Features					
	Split 80:20		5-Fold Validation		LOPO	
	Binary	Multi-class	Binary	Multi-class	Binary	Multi-class
Gradient Boosting	x	x	x	x	x	x
Catboost	x	x	x	x	x	x
LightGBM	x	x	x	x	x	x
SVM	x	x	x	x	x	x
Random Forest	x	x	x	x	x	x
MLP	x					

TABLE XXII: Model investigation with sensor features and top-10 important features. We trained the models with various data split methods: 80/20 train-test split, 5-fold cross-validation, and LOPO cross-validation. The "x" marks in the table indicate the specific model-data combinations investigated in this study.

Model	Sensor Features				Top-10 Important Features			
	5-Fold Validation		LOPO		5-Fold Validation		LOPO	
	Binary	Multi-class	Binary	Multi-class	Binary	Multi-class	Binary	Multi-class
Gradient Boosting	x	x	x	x	x	x	x	x
Catboost	x	x	x	x	x	x	x	x
LightGBM	x	x	x	x	x	x	x	x
SVM	x	x	x	x	x	x	x	x
Random Forest	x	x	x	x	x	x	x	x
MLP	x	x	x	x				

For binary classification, this approach simplifies the prediction task by transforming the DAS levels into a binary classification problem. In particular, we utilize the information from the DAS-21 questionnaire as labels, focusing on the mental health state of the participants. We categorize participants into two classes based on their DAS-21 scores:

- Normal: This class comprises participants who score within the normal range for depression, anxiety, and stress according to established DAS-21 scoring guidelines.
- Abnormal: This class encompasses participants whose DAS-21 scores indicate potential symptoms of depression, anxiety, or stress.

By converting the problem into a binary classification task, we can train models to distinguish between individuals with normal mental health states and those with potential mental health concerns based on their DAS-21 responses and potentially other features from the dataset.

For multi-class classification, this approach aims for a more granular prediction by treating DAS levels as a multi-class classification problem. Instead of collapsing mental health states into two categories, we define multiple classes based on the established DAS-21 scoring ranges: Normal, stress, stress Anxiety, and stress anxiety depression.

B. Data Split

1) Random 80:20 Split: In machine learning, dividing the dataset into training and testing subsets is crucial for evaluating model performance. One common approach is the random 80:20 split, where 80% of the randomly-shuffled data is used for training the model, and the remaining 20% is reserved for testing. This method is favored for its balance between providing enough data for the model to learn effectively and retaining a significant portion for unbiased performance evaluation.

```python
from sklearn.model_selection import train_test_split

# Assuming 'X' is the feature matrix and 'y' is the target vector
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

2) Leave-one-patient-out (LOPO): Predictive models for diagnosis or treatment outcomes must generalize well across different patients. Leave-one-patient-out (LOPO) [30] ensures that the model is evaluated on its ability to perform on new patients not seen during training. For each iteration, all data points from one patient are used as the test set, while the data from all other patients are used to train the model. This process is repeated for each individual in the dataset, ensuring that every patient's data is used as a test set exactly once.

```python
def leave_one_person_out(patient_ids):
  loo = dict()
  for patient_id in set(patient_ids):
    loo[patient_id] = None

  for patient_id in loo:
    train_indices, test_indices = [], []

    for i in range(len(patient_ids)):
      if patient_ids[i] == patient_id:
        test_indices.append(i)
      else:
        train_indices.append(i)

    loo[patient_id] = [train_indices, test_indices]

  return loo


patient_ids = list(df_rawdata['Patient_ID'])
loo = leave_one_person_out(patient_ids)


for patient_id in loo:
  # Leave one out
  train_indices = np.array(loo[patient_id][0])
  test_indices = np.array(loo[patient_id][1])

  X_train = X[train_indices]
  X_test = X[test_indices]
  y_train = y[train_indices]
  y_test = y[test_indices]
```

3) K-Folds: K-fold cross-validation is a robust technique for training machine learning models in healthcare, ensuring reliable and unbiased performance evaluation [30]. K-fold cross-validation mitigates the risk of overfitting, maximizes data utilization, and enhances the model's generalizability, which is crucial for accurately predicting patient outcomes and making informed medical decisions. It involves randomly partitioning the dataset into k equally sized folds, where the model is trained on k-1 folds and validated on the remaining fold. This process is repeated k times, with each fold serving as the validation set once, and the performance metrics are averaged to provide a comprehensive assessment. Note that for each fold there is no patient overlapping. In the scope of this work, we used 5-fold cross-validation and 10 seeds for training, which means 5-fold was trained for 10 times to avoid varying results.

```python
def k_fold(patient_ids, number_of_folds = 5):
  patient_id_dict = dict()
  for patient_id in set(patient_ids):
    patient_id_dict[patient_id] = []

  for i in range(len(patient_ids)):
    patient_id = patient_ids[i]
    patient_id_dict[patient_id].append(i)

  kfold_dict = dict()
  i, k = 0, len(patient_id_dict.keys())//number_of_folds
  queue = list(patient_id_dict.keys())

  while queue:
    random.shuffle(queue)
    patient_id_test = [j for j in queue[0:k] ]
    patient_id_train = [j for j in set(patient_id_dict.keys())-set(patient_id_test) ]

    train_indices, test_indices = [], []
    for patient_id in patient_id_train:
      train_indices += patient_id_dict[patient_id]
    for patient_id in patient_id_test:
      test_indices += patient_id_dict[patient_id]
```

```
        kfold_dict[i] = [train_indices, test_indices]

        queue = queue[k:]
        i += 1

    return kfold_dict


patient_ids = list(df_rawdata['Patient_ID'])


for seed in range(10):
    kfold = k_fold(patient_ids, number_of_folds=5)
    for k in kfold:
        # K-fold
        train_indices = np.array(kfold[k][0])
        test_indices = np.array(kfold[k][1])

        X_train = X[train_indices]
        X_test = X[test_indices]
        y_train = y[train_indices]
        y_test = y[test_indices]
```

C. Evaluation Metric

To assess the performance of the machine learning models for predicting Depression, Anxiety, and stress (DAS) levels, we employ two key evaluation metrics: Receiver Operating Characteristic (ROC) AUC (Area Under the Curve) and Precision-Recall (PR) AUC. These metrics provide a comprehensive assessment of the model's discriminative ability and its performance in handling class imbalances.

The ROC AUC metric measures the ability of the model to distinguish between classes. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The AUC value ranges from 0 to 1, where a value of 1 indicates perfect classification and a value of 0.5 suggests performance no better than random guessing. A higher ROC AUC indicates that the model is better at differentiating between the positive and negative classes.

In our study, we calculate the ROC AUC to assess how well our model can predict DAS levels across the entire range of possible thresholds. This metric is particularly useful in healthcare applications, where the cost of false positives and false negatives can vary, and understanding the trade-offs between sensitivity and specificity is crucial.

While ROC AUC provides a good overall measure of performance, it can be insensitive to class imbalance. In our case, the prevalence of depression, anxiety, and stress might be lower than the prevalence of healthy individuals. The PR curve addresses this by plotting Precision against Recall.

This section shows the ROC AUC and ROC PR plots for all binary classification results above.

A. All Features with 80:20 Split (Binary Classification)

Figure 9, 10, 11, 12, 13, 14 shows the ROC AUC and ROC PR plots for all features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a random 80:20 split. These figures are plots from Table IV in Section IV-A.
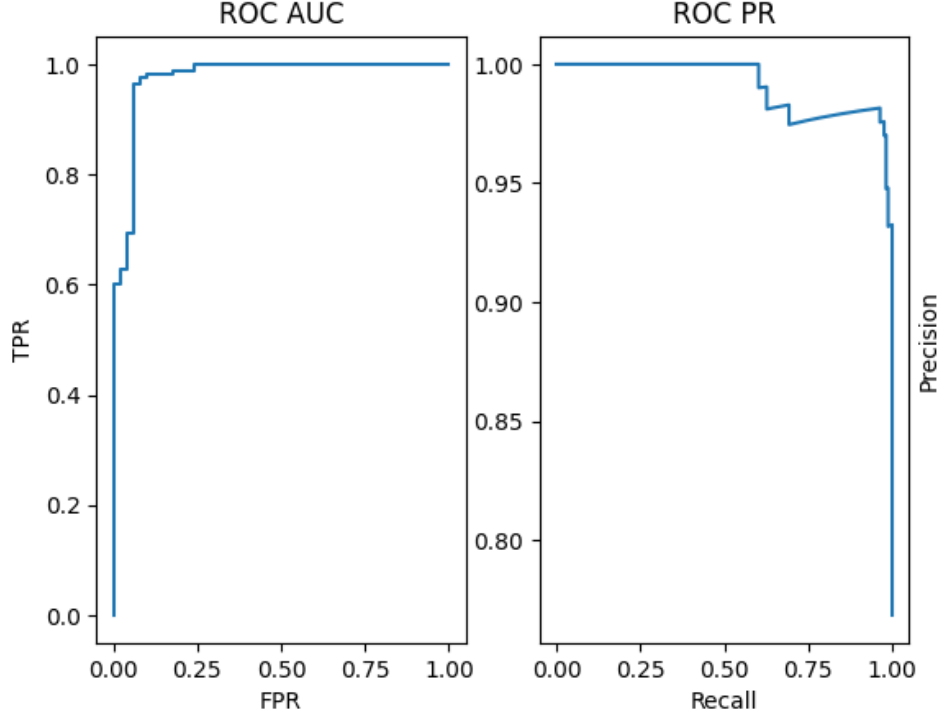
Fig. 9: All features with 80:20 split, binary classification: Gradient Boosting

B. All Features with Cross-Validation

1) Binary classification: LOPO: Figure 15, 16, 17, 18, 19, 20 shows the ROC AUC and ROC PR plots for all features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a LOPO split. These figures are plots from Table VIII in Section IV-B.

2) Binary classification: K-folds: Figure 21, 22, 23, 24, 25, 26 shows the ROC AUC and ROC PR plots for all features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a K-folds split. These figures are plots from Table IX in Section IV-B.

C. Using multimodal sensor features classification

1) Binary classification: LOPO: Figure 27, 28, 29, 30, 31, 32 shows the ROC AUC and ROC PR plots for multimodal sensor features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a LOPO split. These figures are plots from Table XII in Section IV-C.

2) Binary classification: K-folds: Figure 33, 34, 35, 36, 37, 38 shows the ROC AUC and ROC PR plots for multimodal sensor features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a K-folds split. These figures are plots from Table XIII in Section IV-C.

D. Using Top-10 important features classification

1) Binary classification: LOPO: Figure 39, 40, 41, 42, 43, 44 shows the ROC AUC and ROC PR plots for top 10 features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a LOPO split. These figures are plots from Table XIV in Section IV-D.
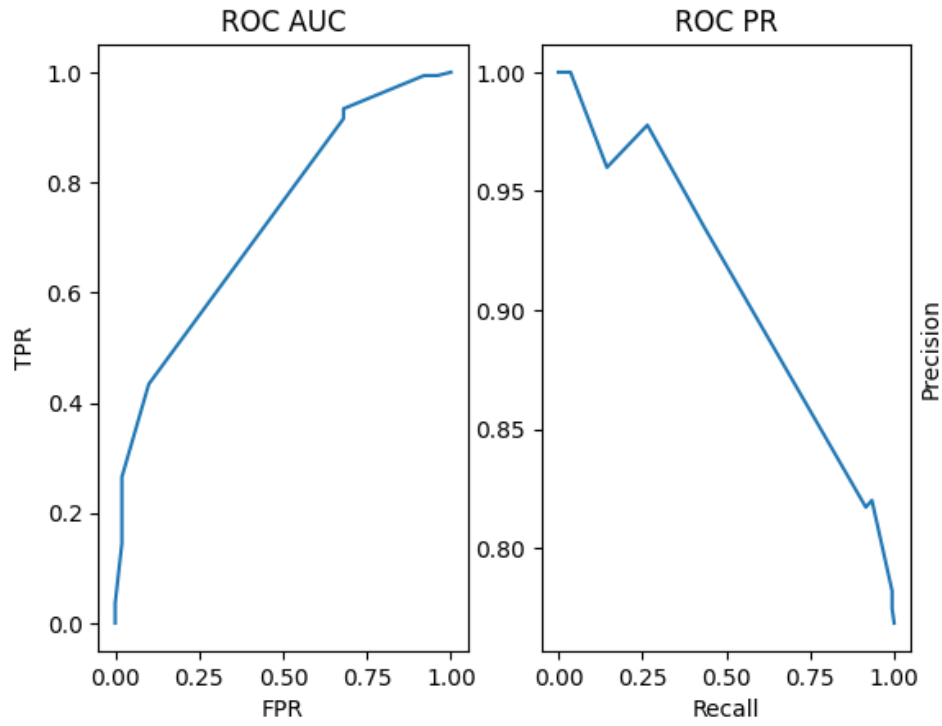
Fig. 10: All features with 80:20 split, binary classification: Catboost

2) Binary classification: K-folds: Figure 45, 46, 47, 48, 49 shows the ROC AUC and ROC PR plots for top 10 features binary classification models using Gradient Boosting, Catboost, LightGBM, SVM, Random Forest, and MLP respectively, with a K-folds split. These figures are plots from Table XV in Section IV-D.
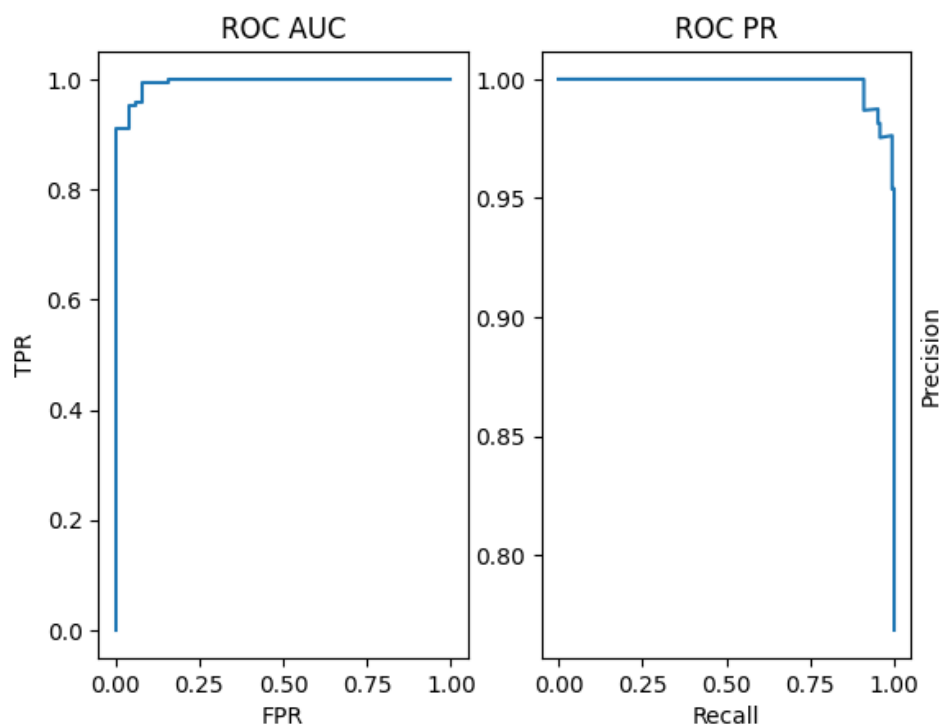
Fig. 11: All features with 80:20 split, binary classification: LightGBM
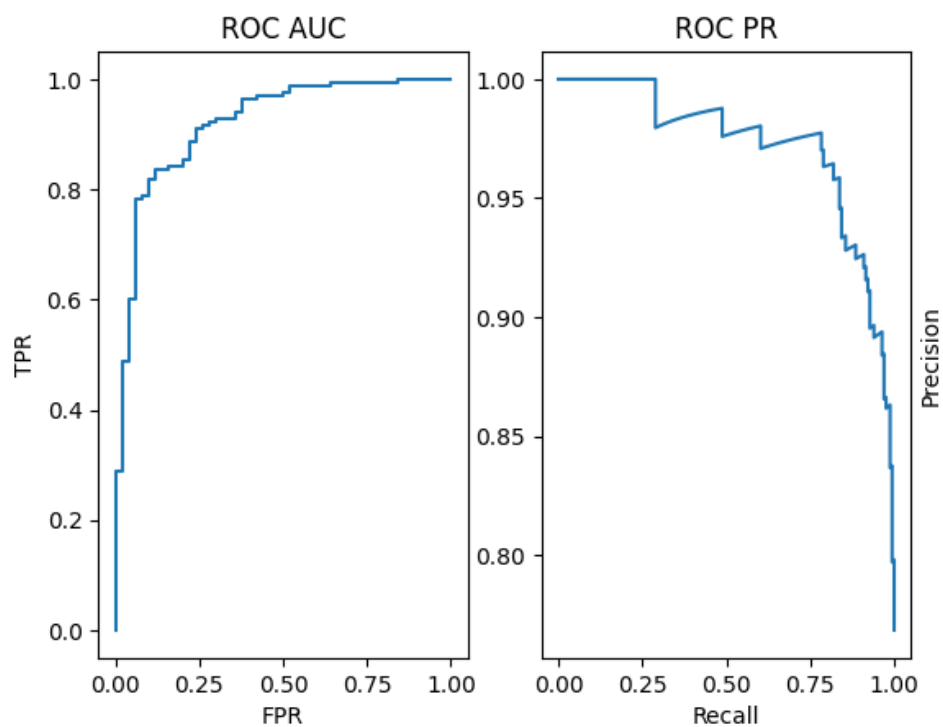
Fig. 12: All features with 80:20 split, binary classification: SVM
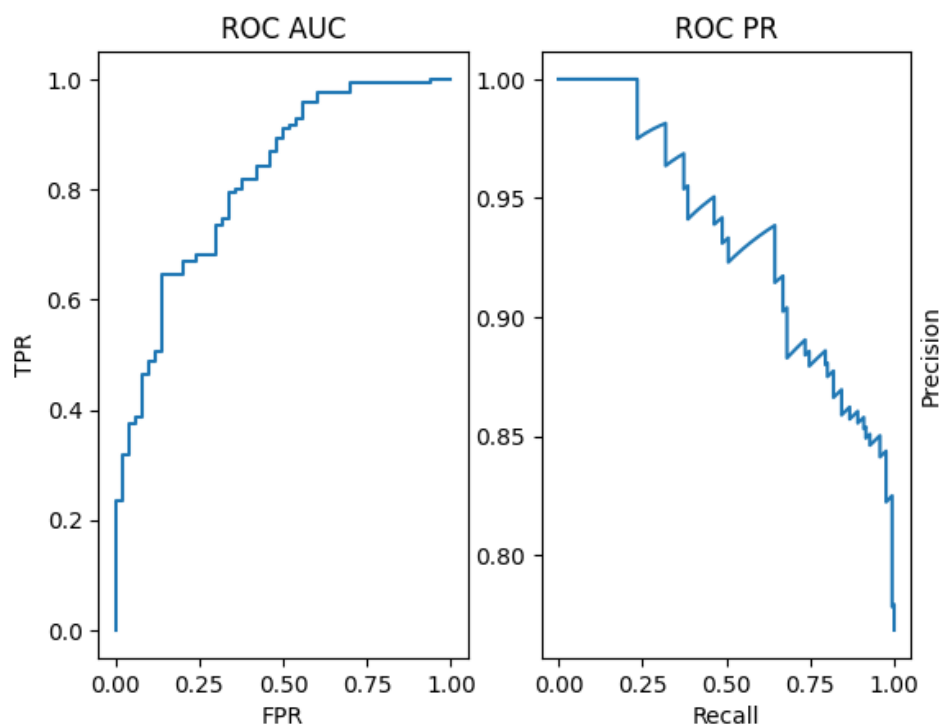
Fig. 13: All features with 80:20 split, binary classification: Random forest
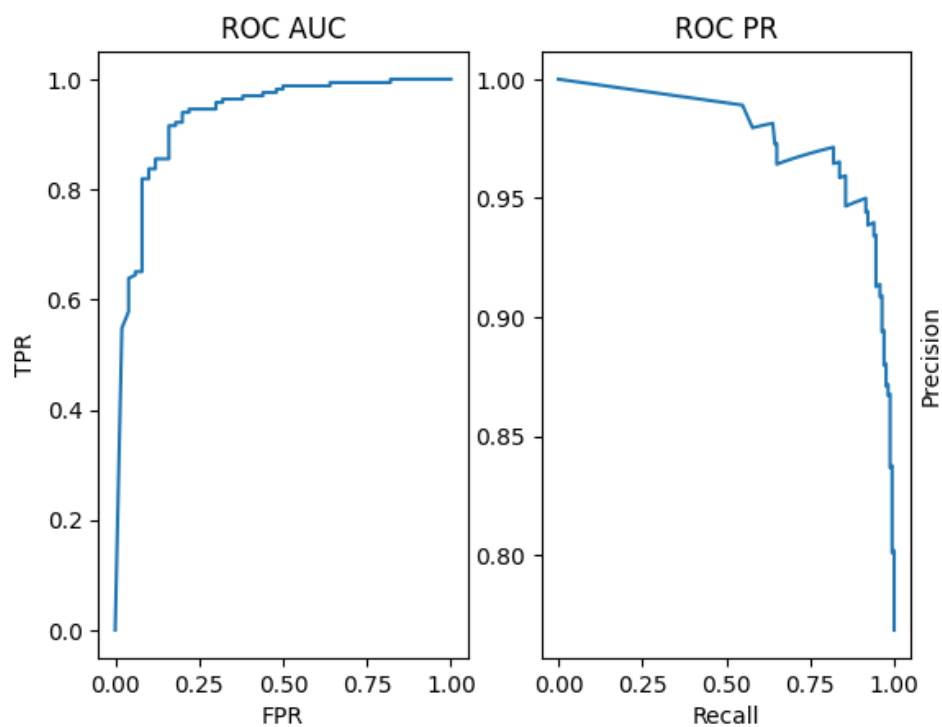
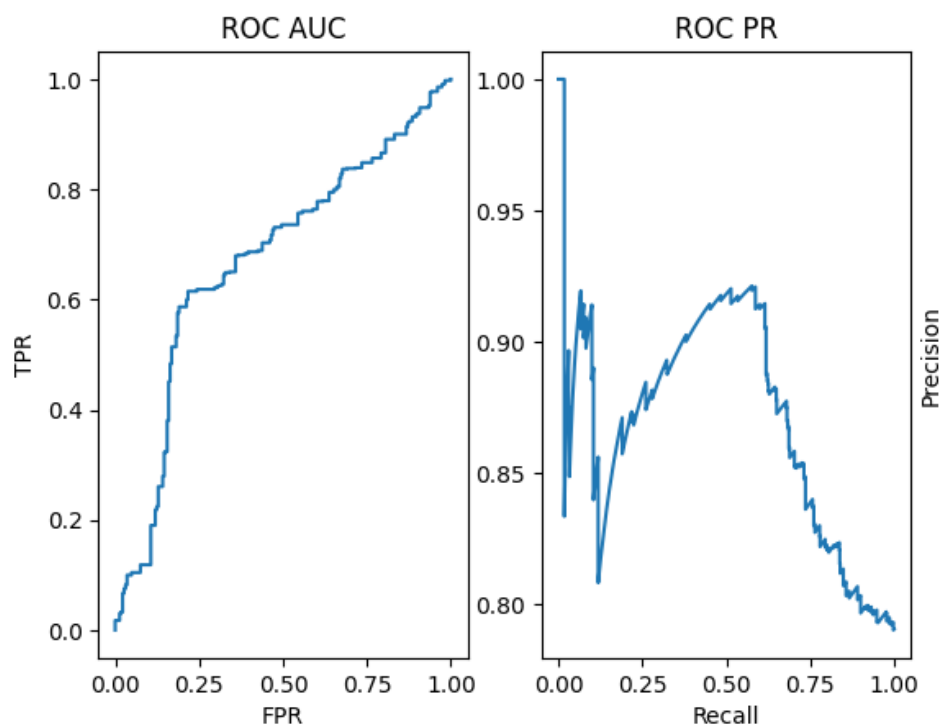Fig. 14: All features with 80:20 split, binary classification: MLP
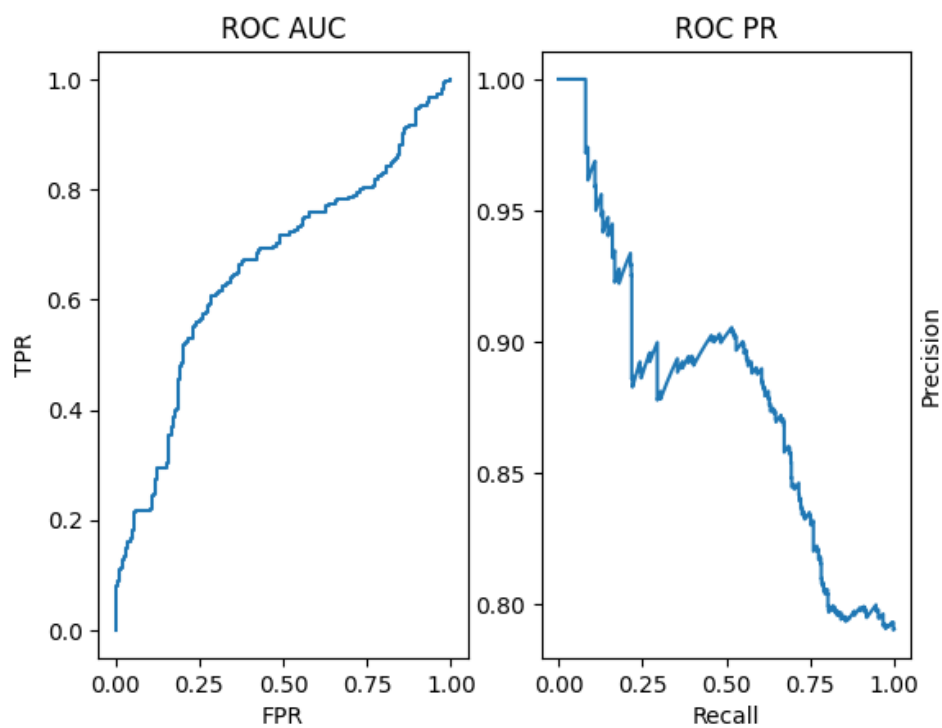
Fig. 15: All features with LOPO, binary classification: Gradient Boosting
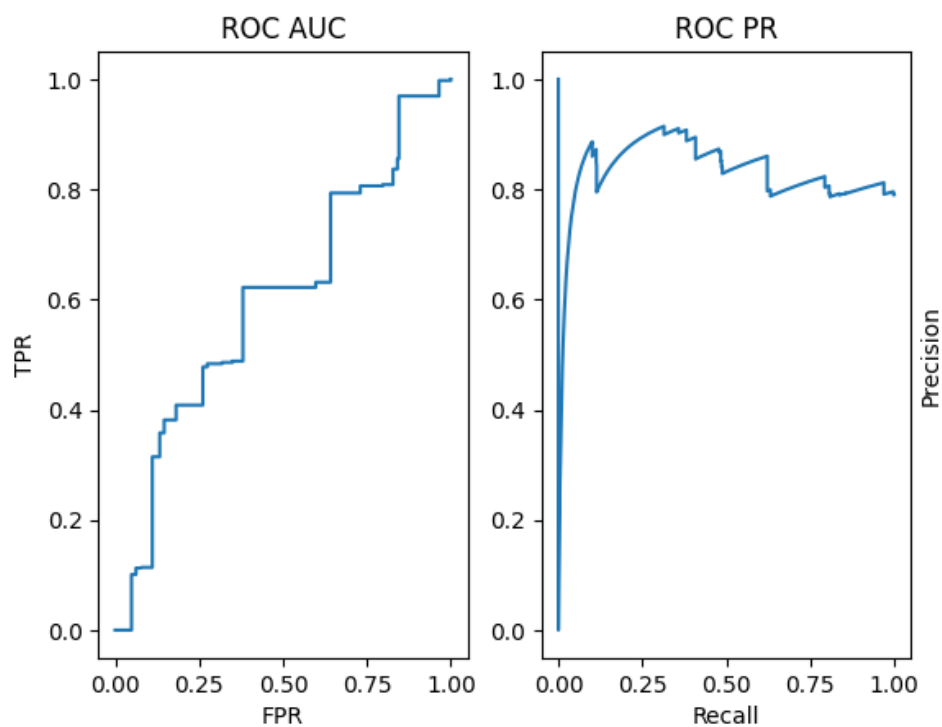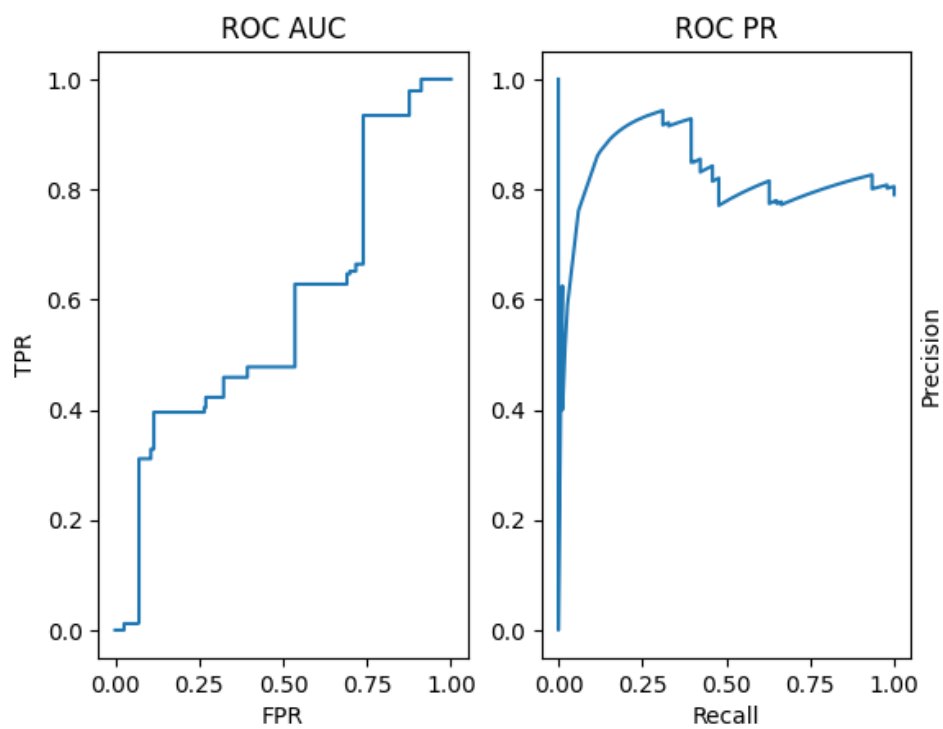

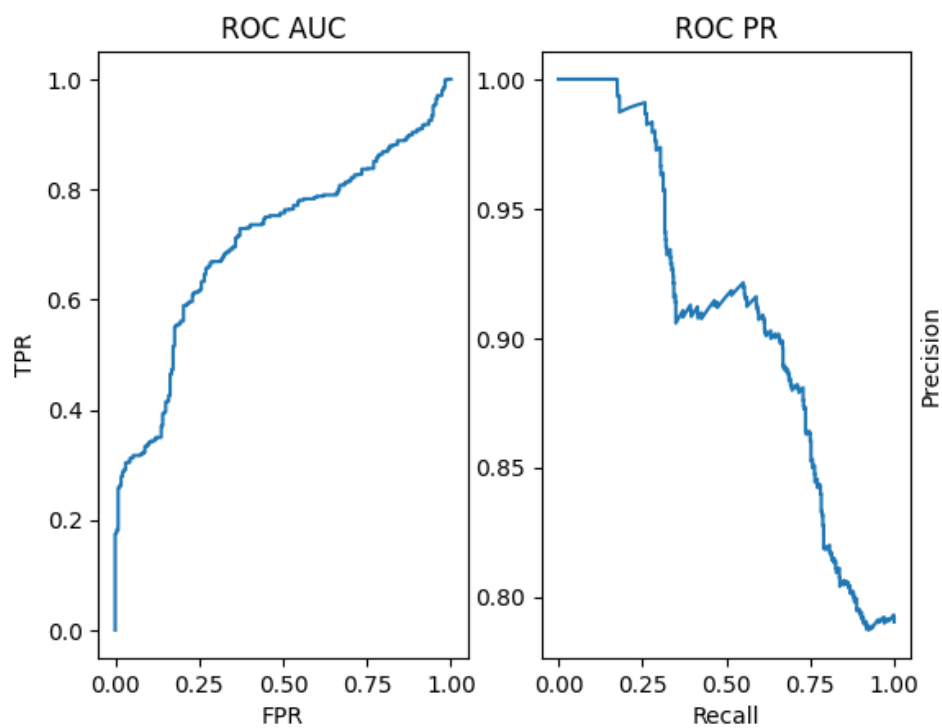Fig. 16: All features with LOPO, binary classification: Catboost
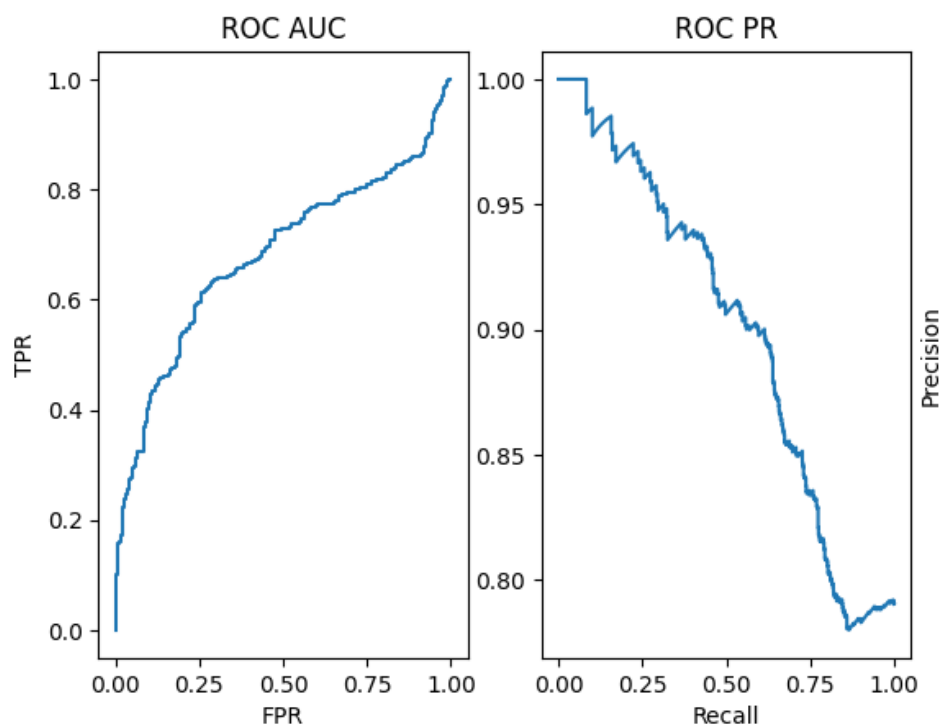
Fig. 17: All features with LOPO, binary classification: LightGBM
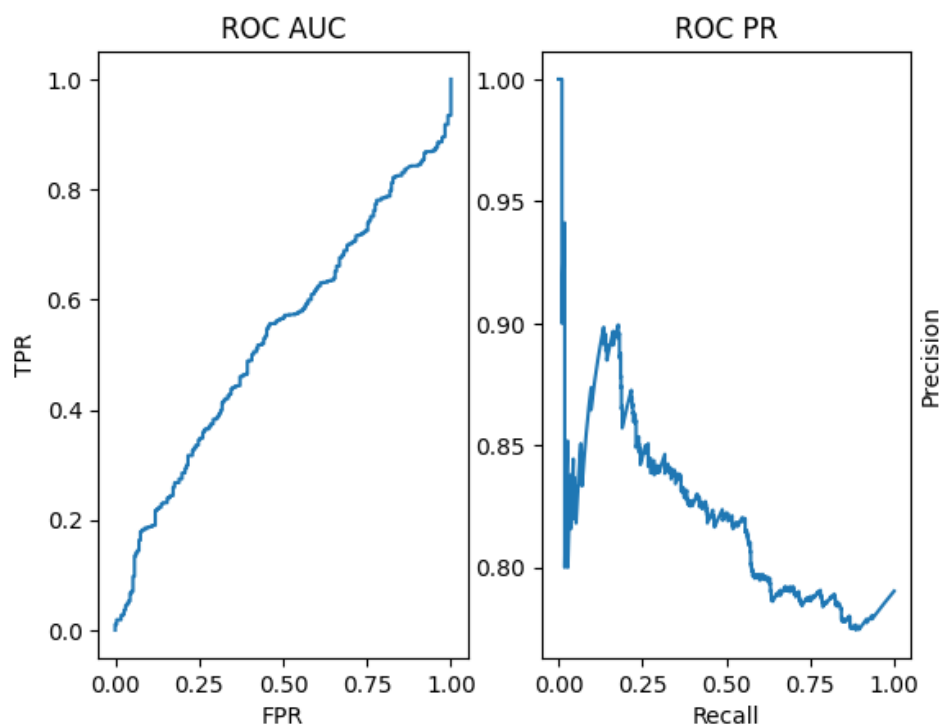
Fig. 18: All features with LOPO, binary classification: SVM
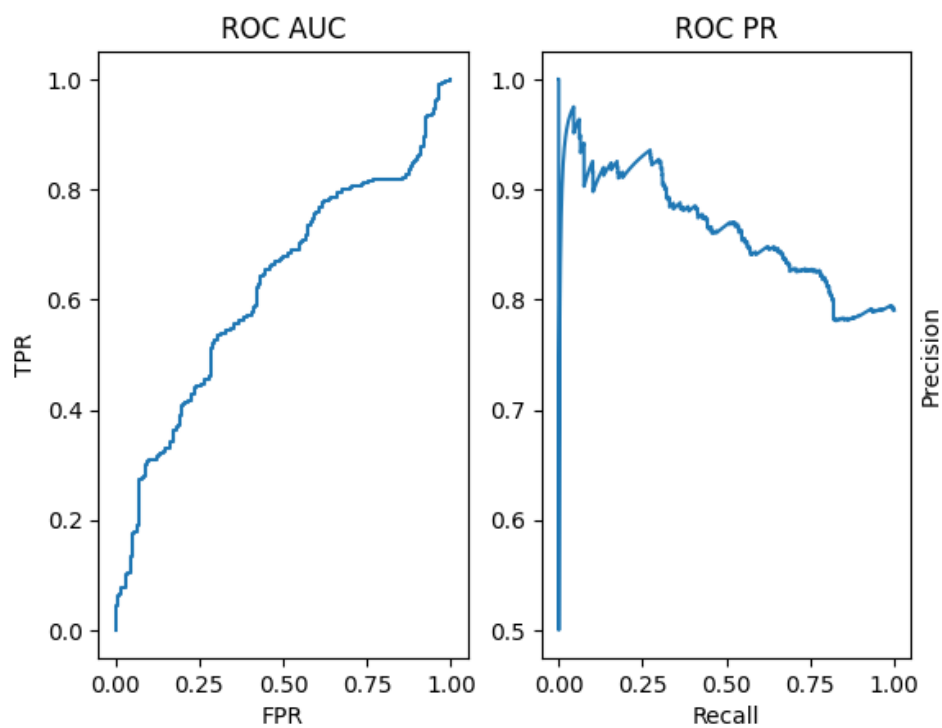
Fig. 19: All features with LOPO, binary classification: Random Forest
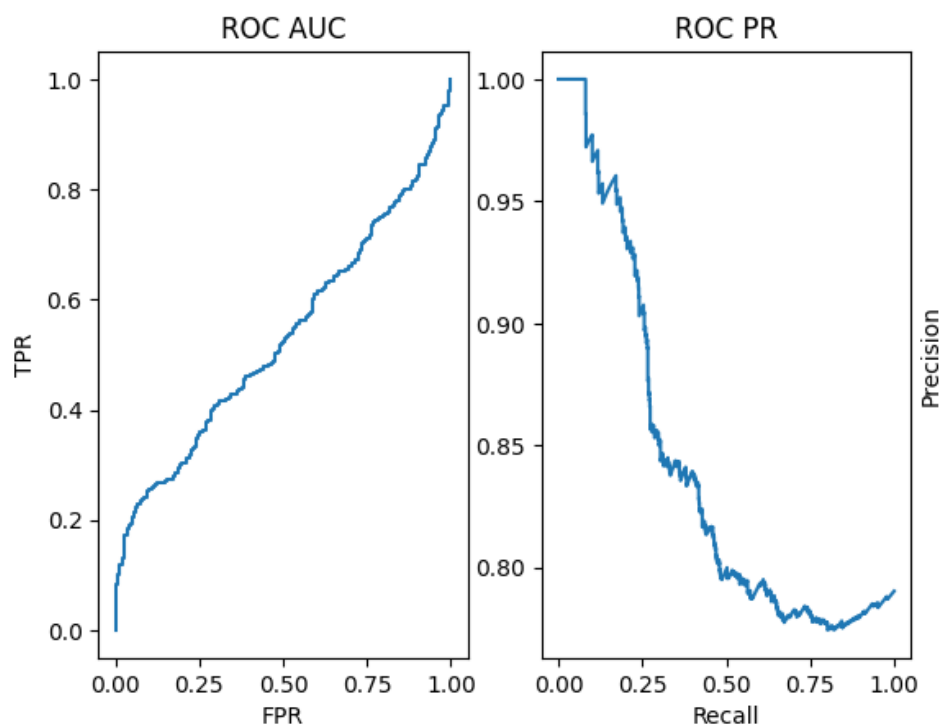
Fig. 20: All features with LOPO, binary classification: MLP
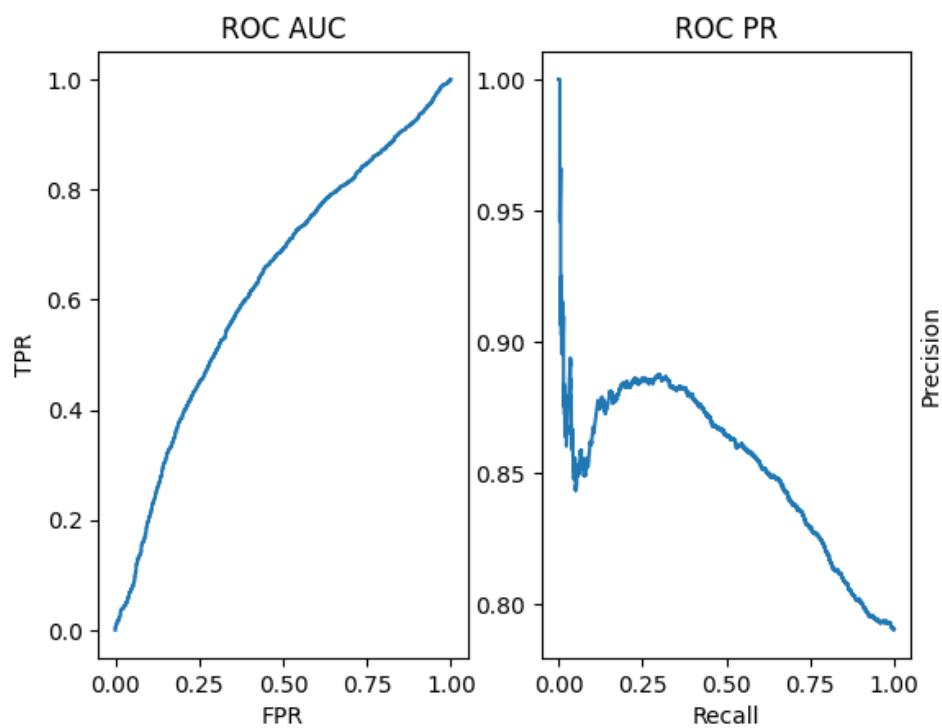
Fig. 21: All features with K-folds, binary classification: Gradient Boosting
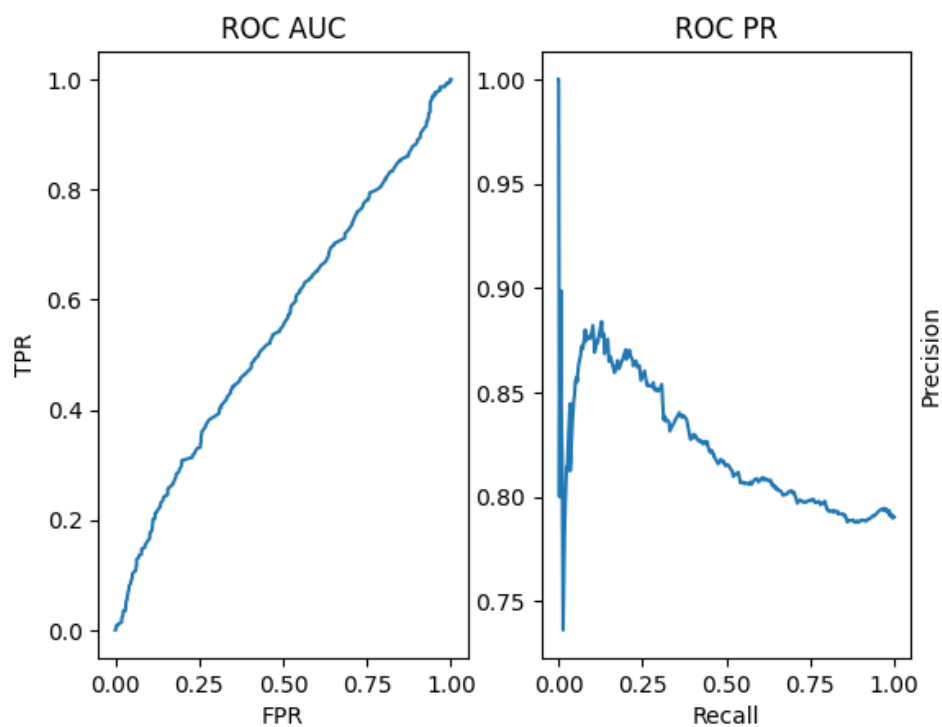
Fig. 22: All features with K-folds, binary classification: Catboost
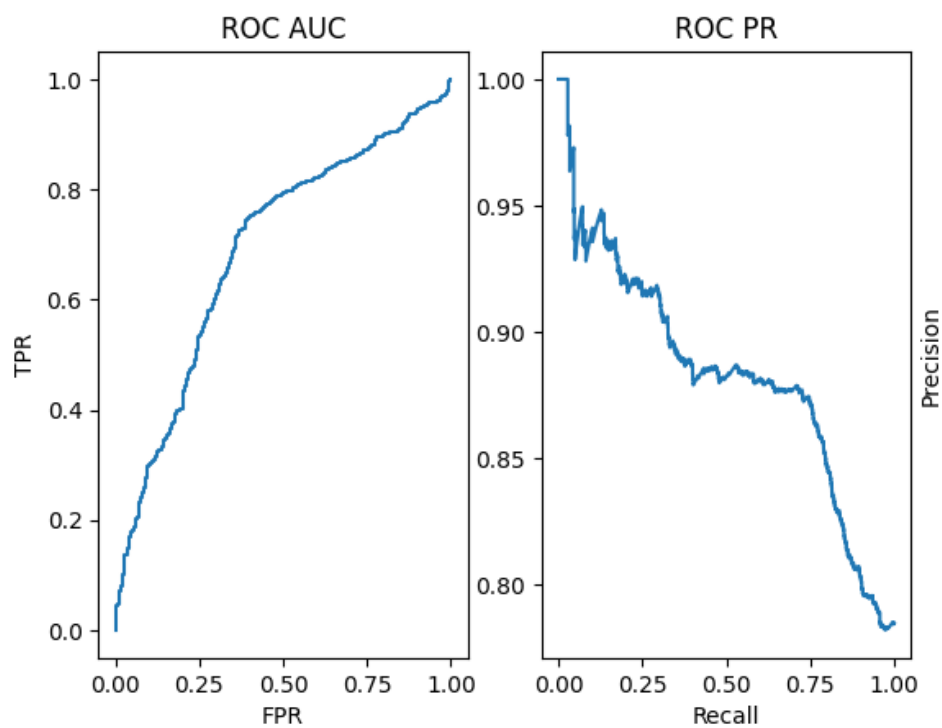
Fig. 23: All features with K-folds, binary classification: LightGBM
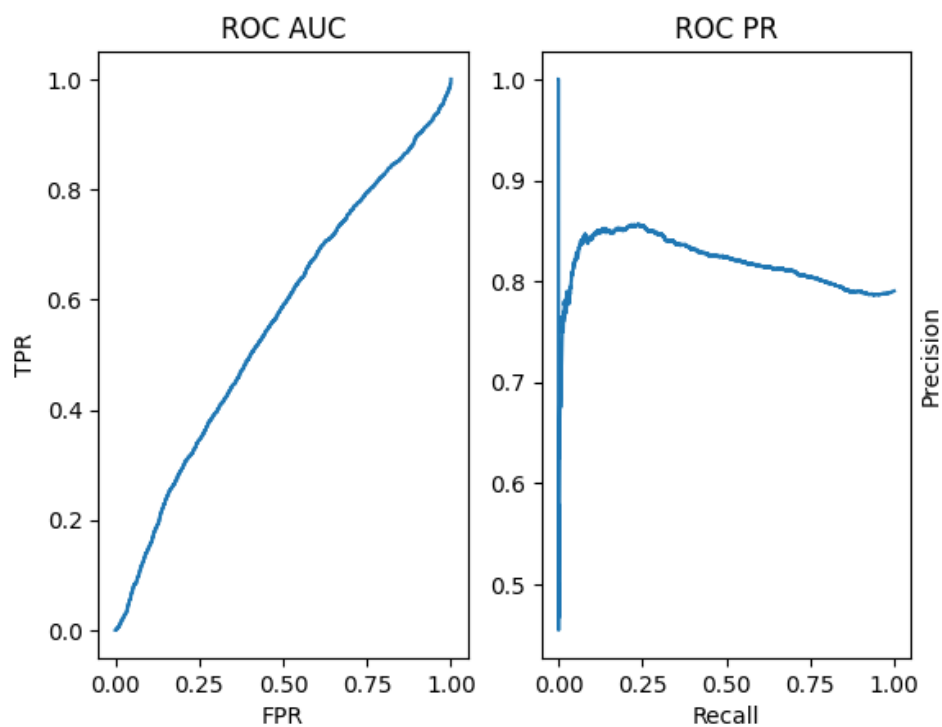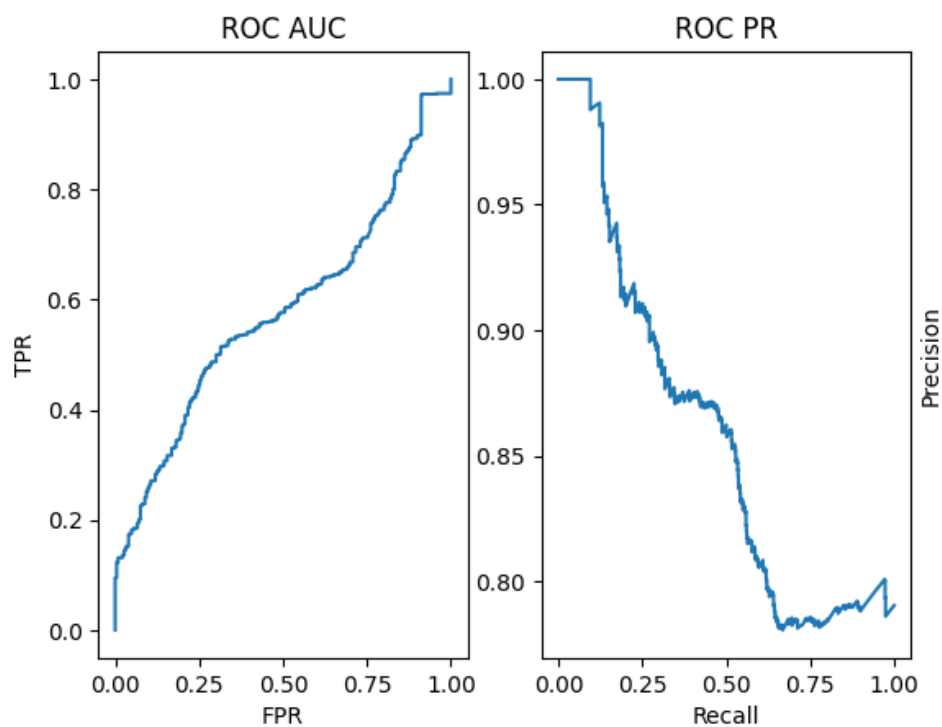

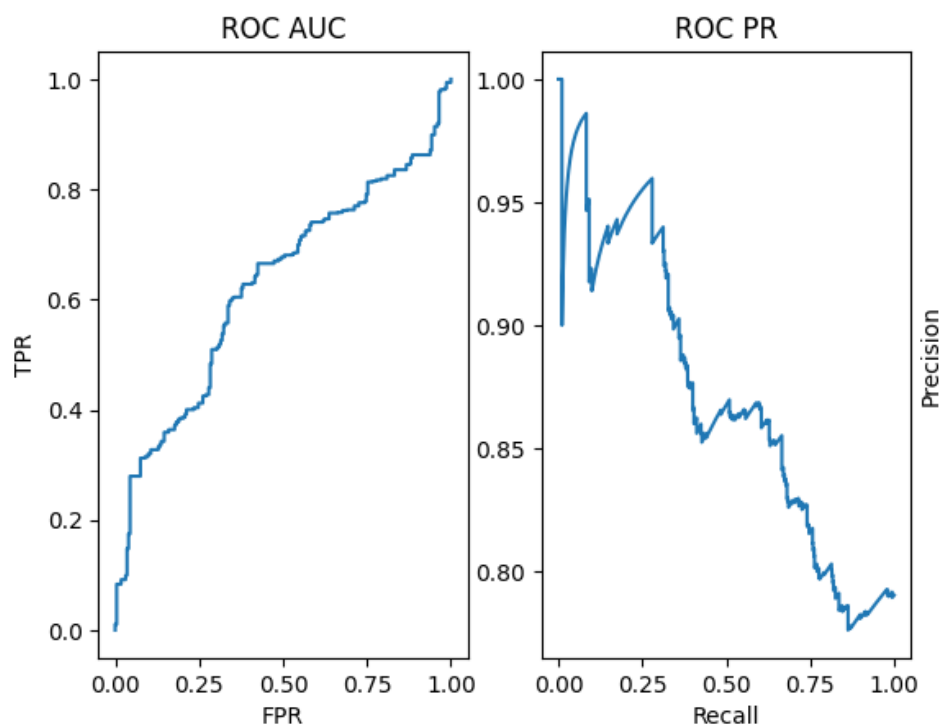Fig. 24: All features with K-folds, binary classification: SVM
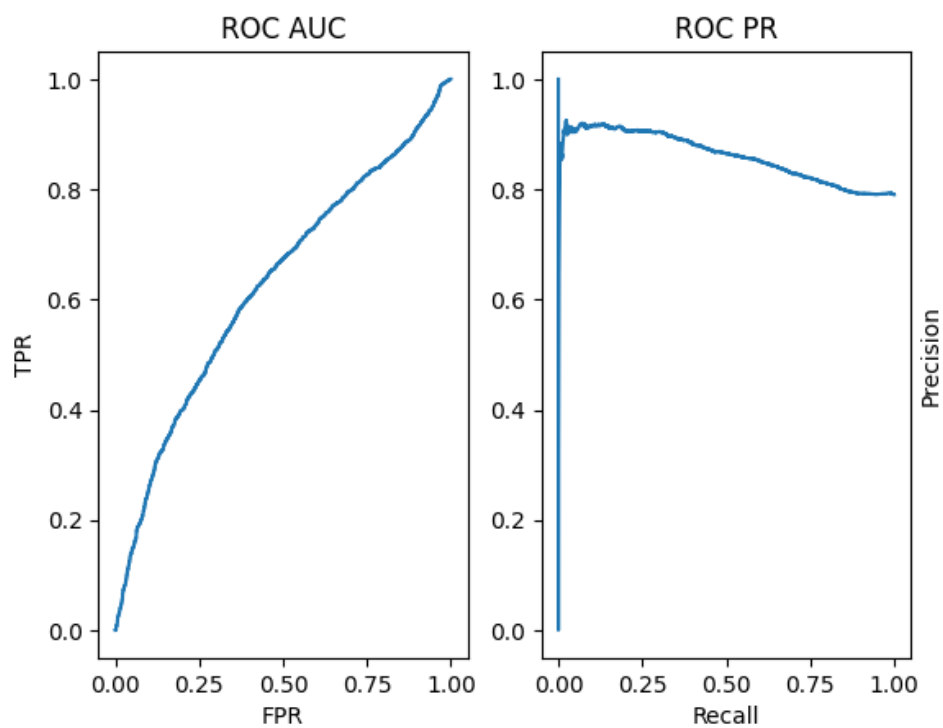
Fig. 25: All features with K-folds, binary classification: Random Forest
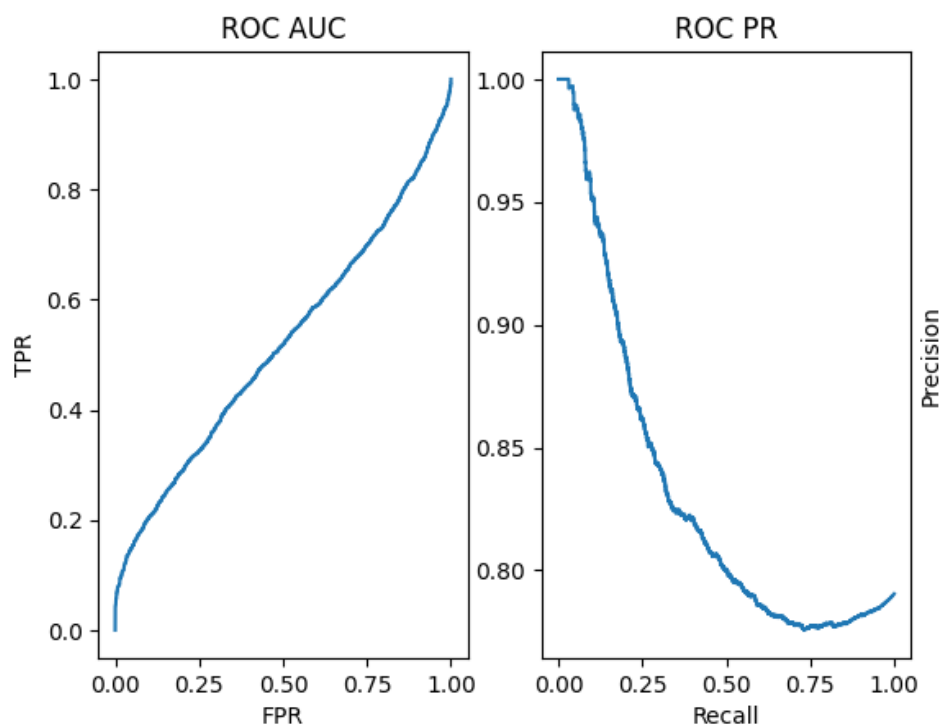
Fig. 26: All features with K-folds, binary classification: MLP
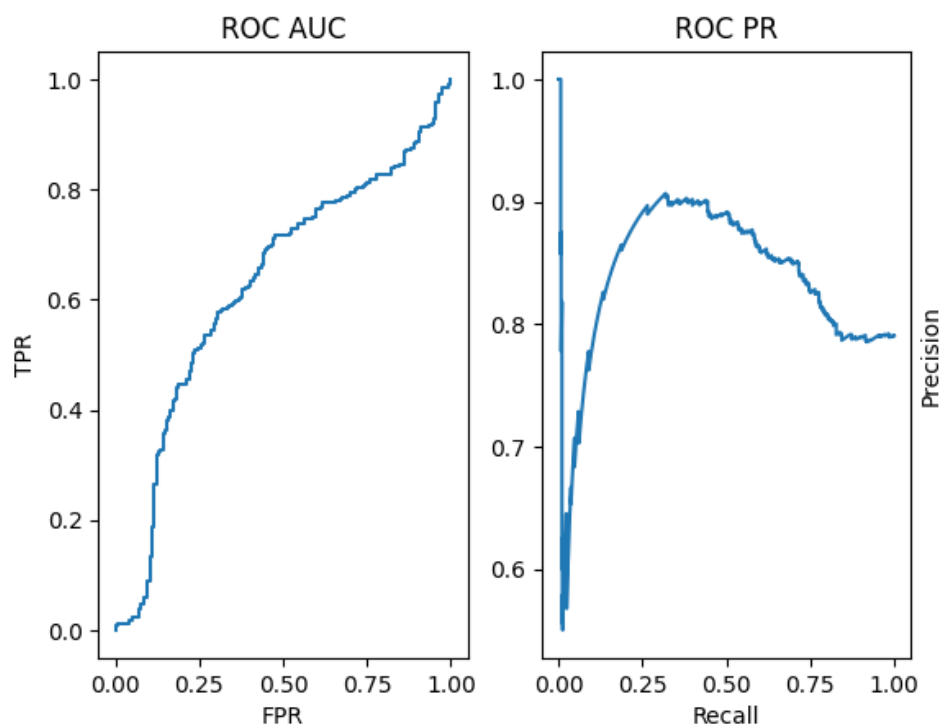
Fig. 27: Multimodal sensor features with LOPO, binary classification: Gradient Boosting
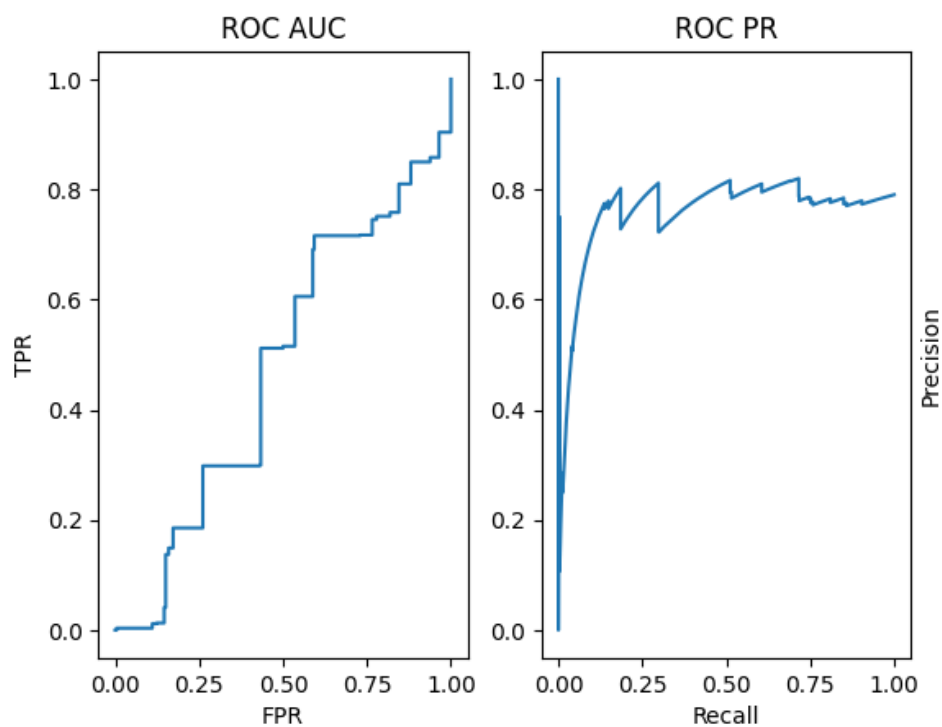
Fig. 28: Multimodal sensor features with LOPO, binary classification: Catboost
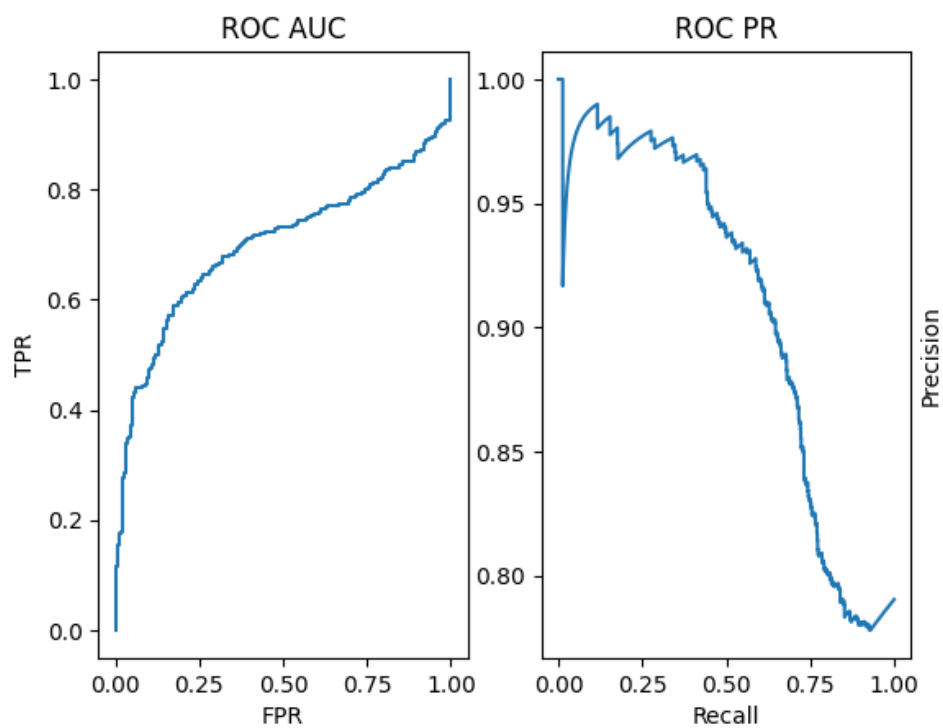
Fig. 29: Multimodal sensor features with LOPO, binary classification: LightGBM
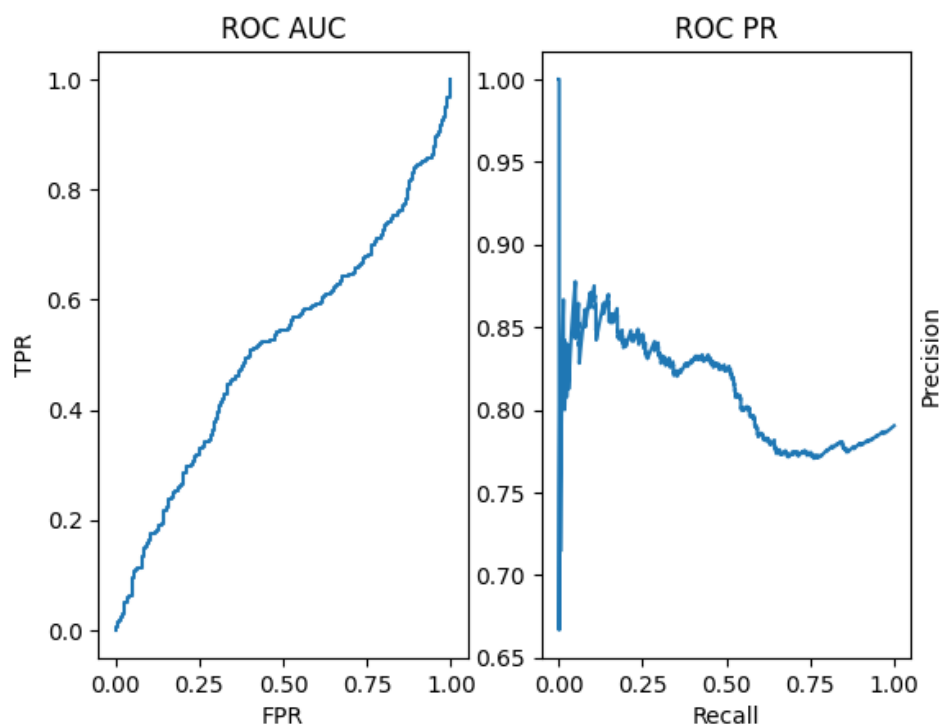
Fig. 30: Multimodal sensor features with LOPO, binary classification: SVM
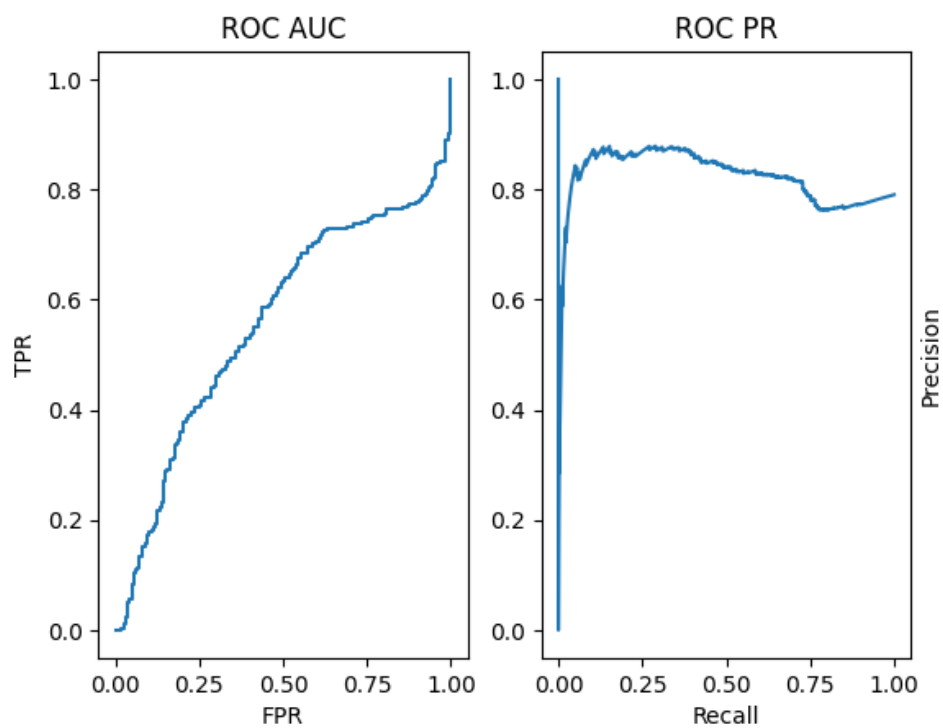
Fig. 31: Multimodal sensor features with LOPO, binary classification: Random Forest
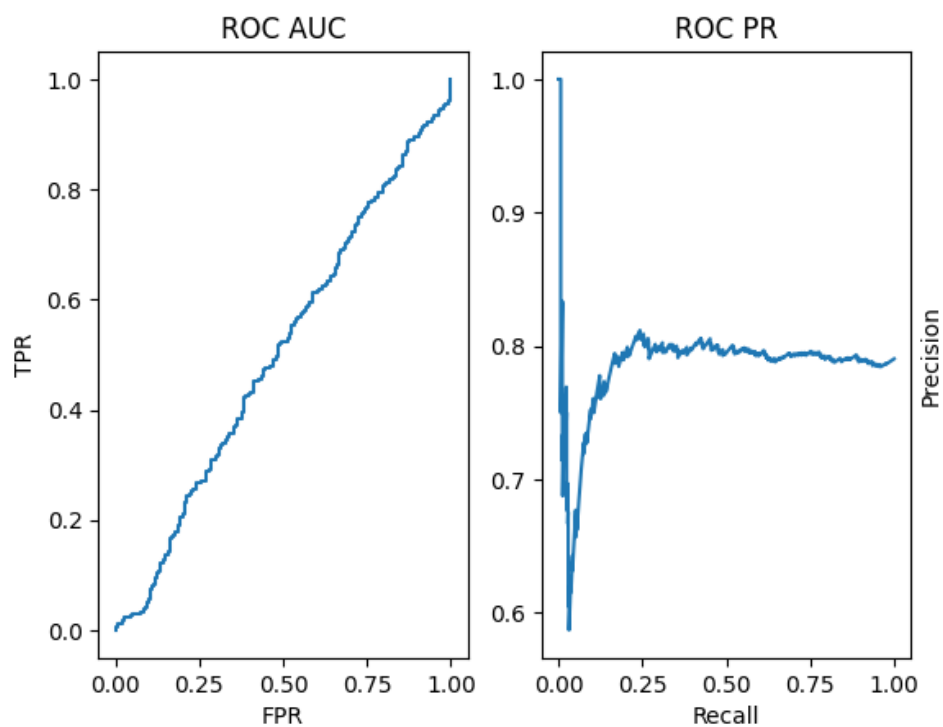
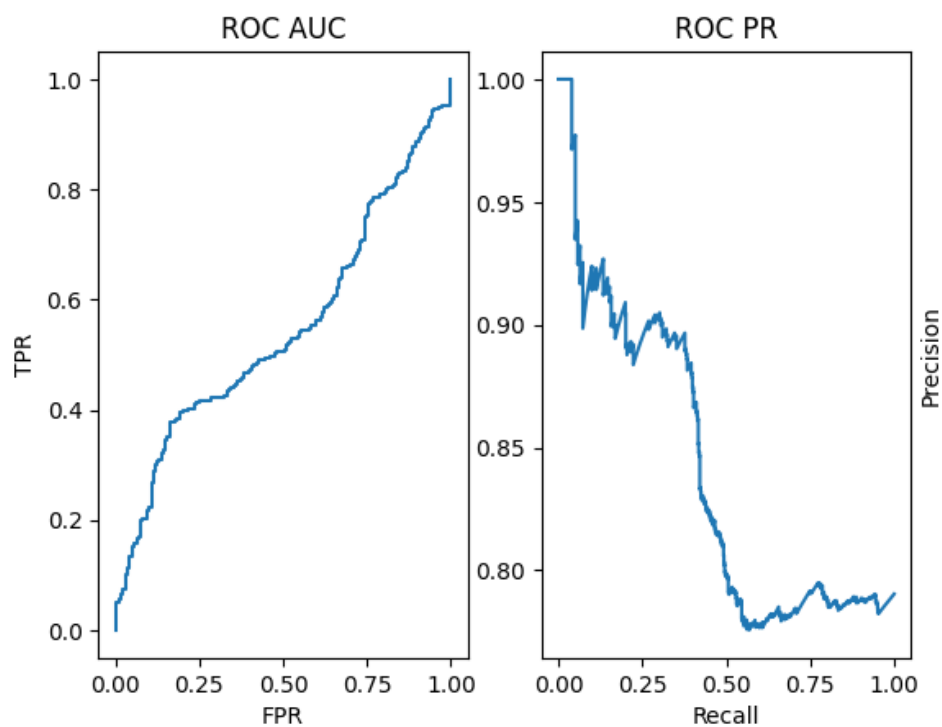Fig. 32: Multimodal sensor features with LOPO, binary classification: MLP
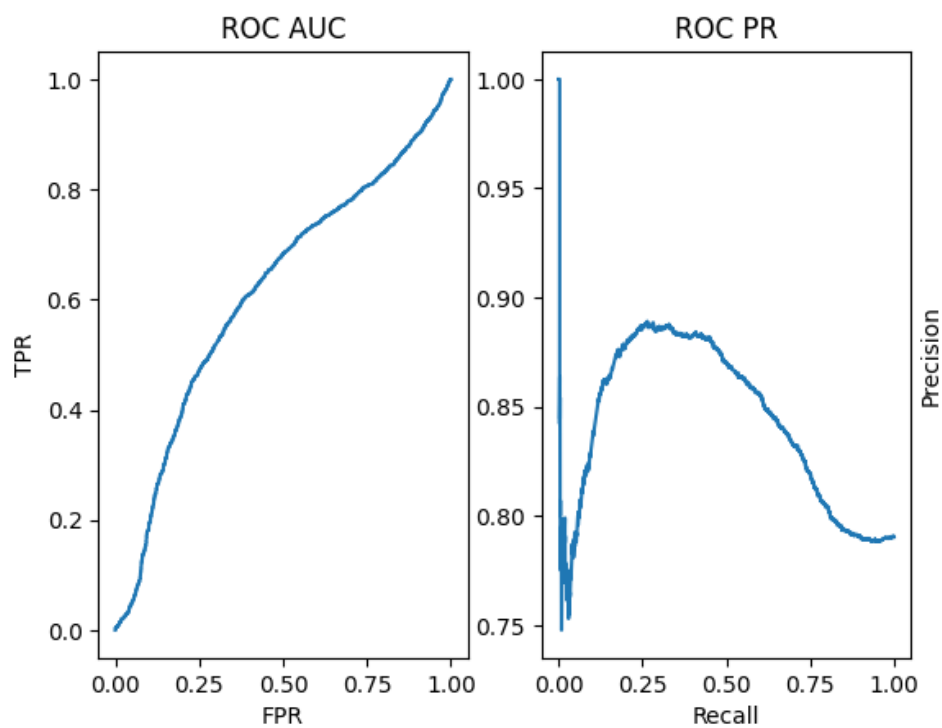
Fig. 33: Multimodal sensor features with K-folds, binary classification: Gradient Boosting
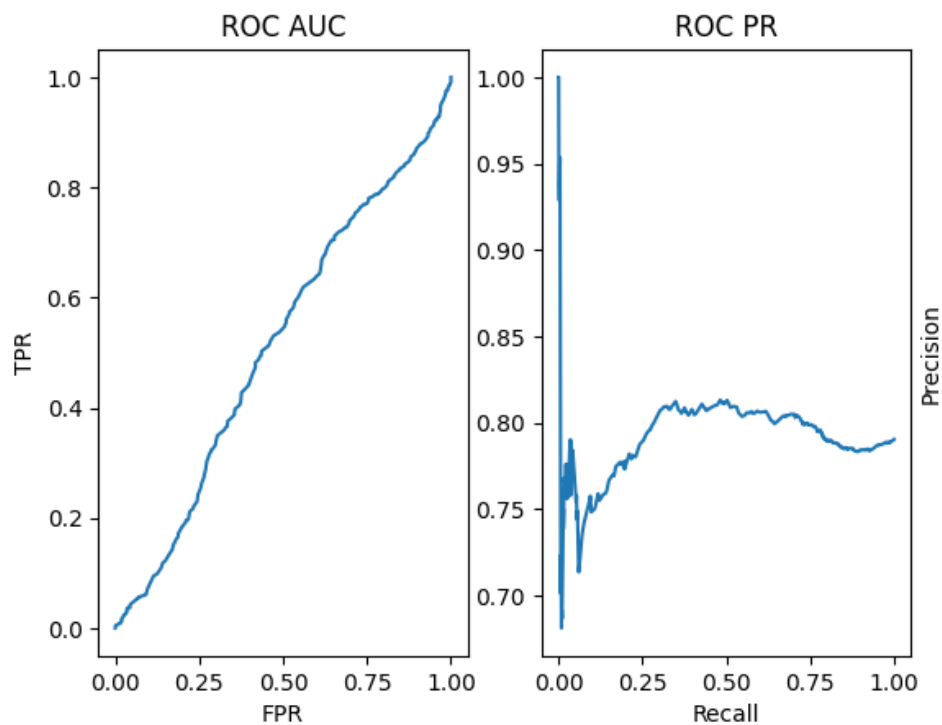
Fig. 34: Multimodal sensor features with K-folds, binary classification: Catboost
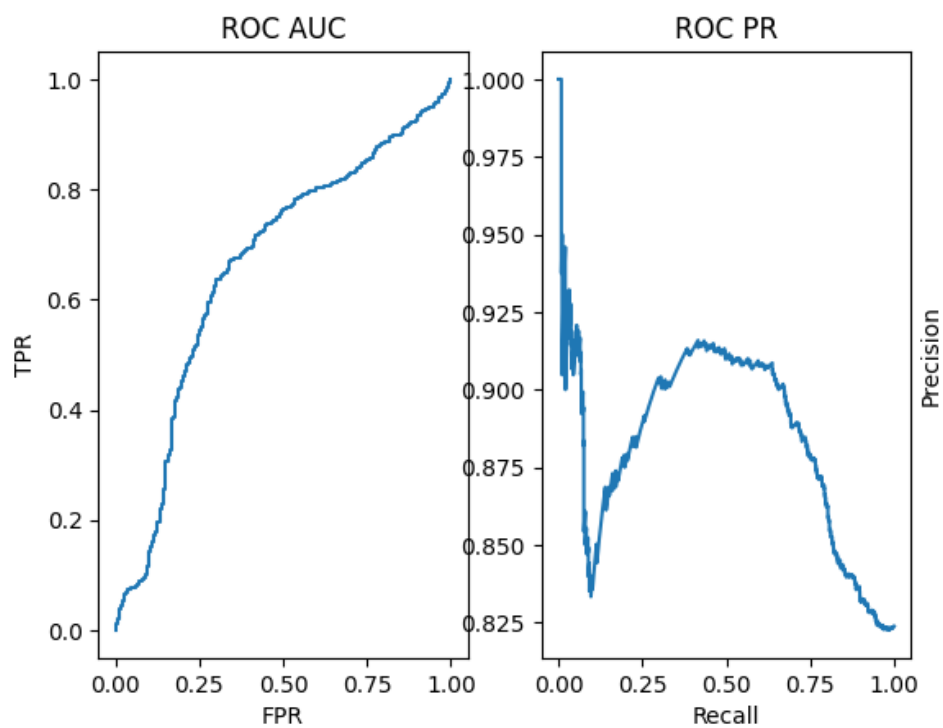
Fig. 35: Multimodal sensor features with K-folds, binary classification: LightGBM
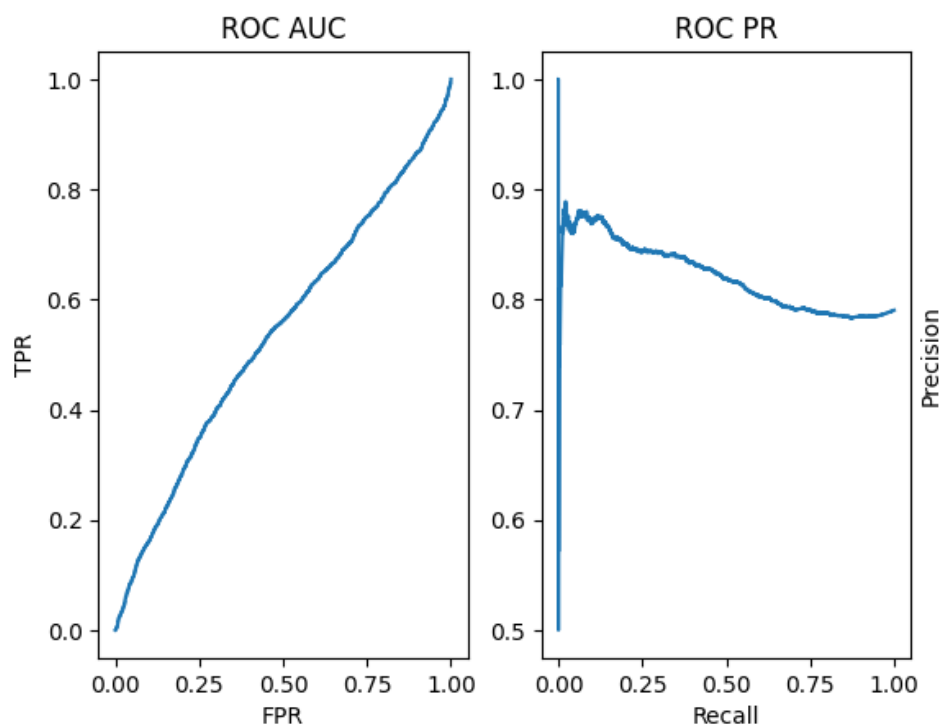
Fig. 36: Multimodal sensor features with K-folds, binary classification: SVM
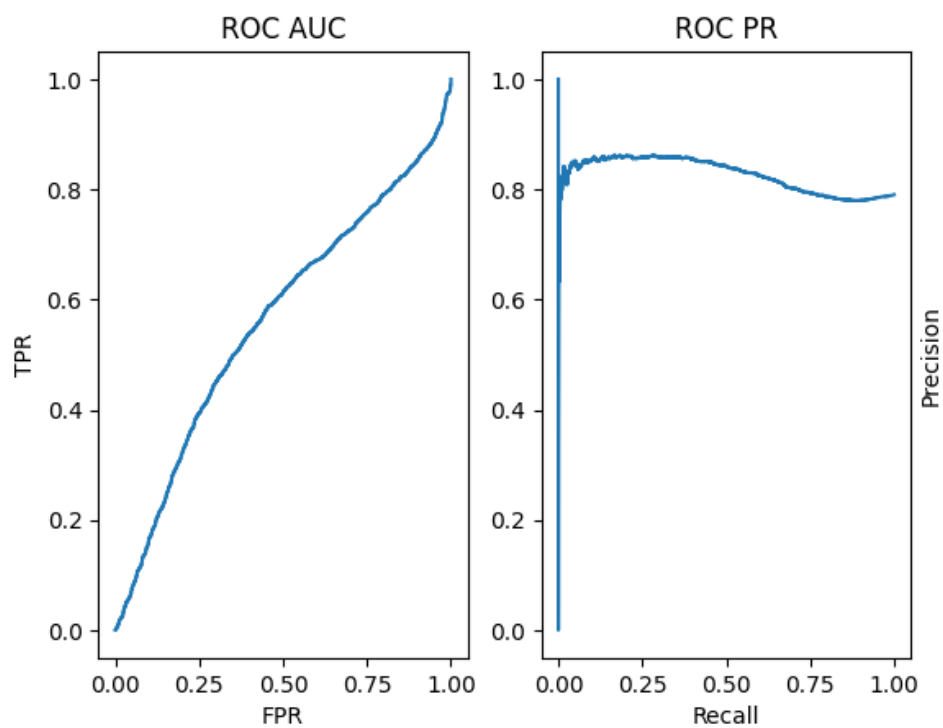
Fig. 37: Multimodal sensor features with K-folds, binary classification: Random Forest
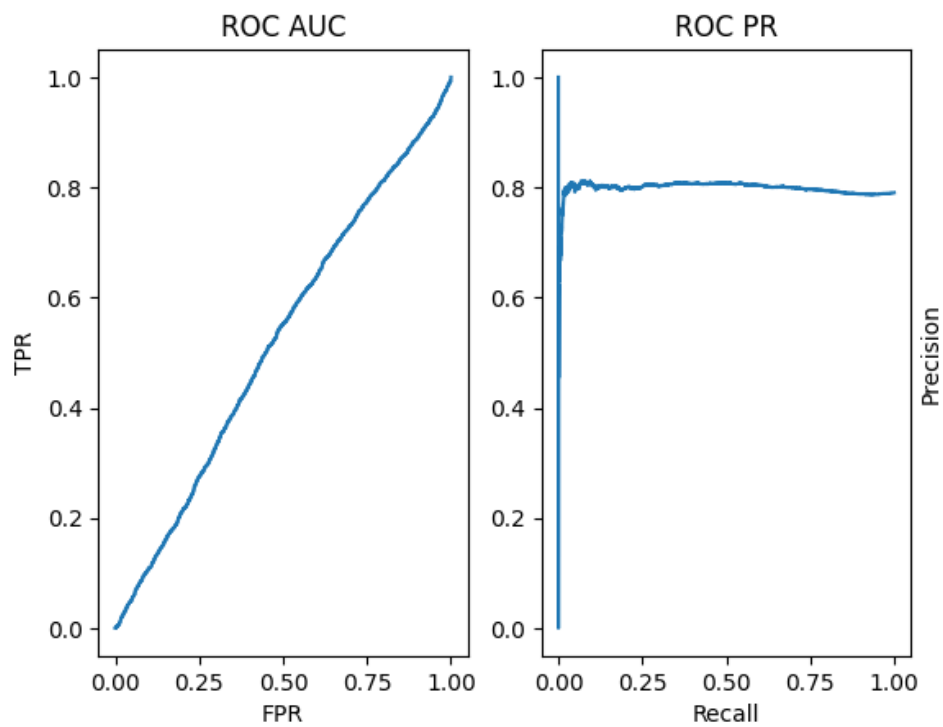
Fig. 38: Multimodal sensor features with K-folds, binary classification: MLP

Fig. 39: Top 10 features with LOPO, binary classification: Gradient Boosting

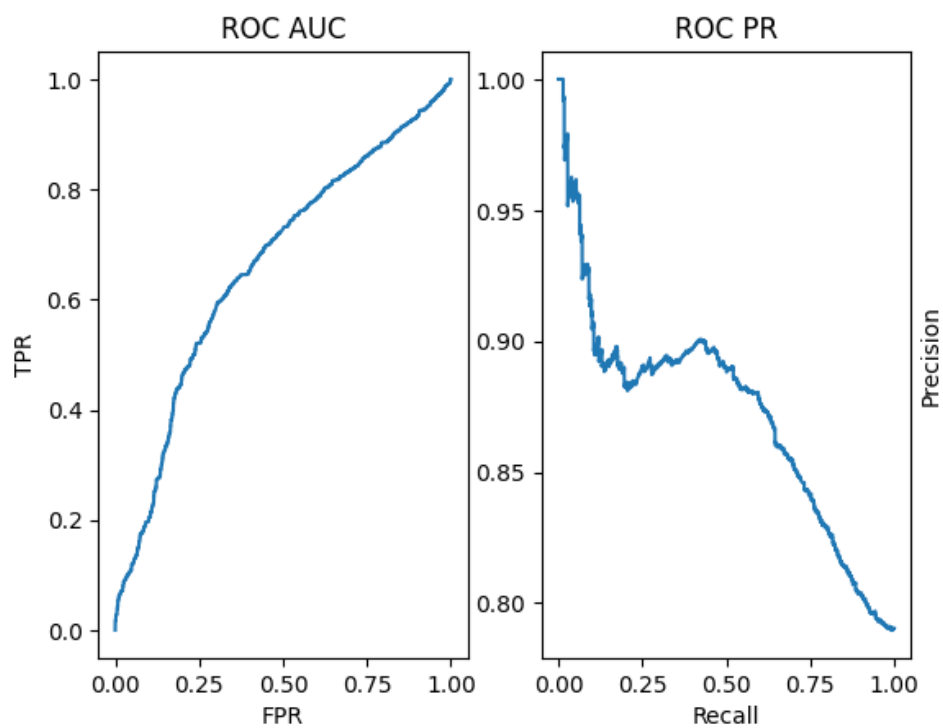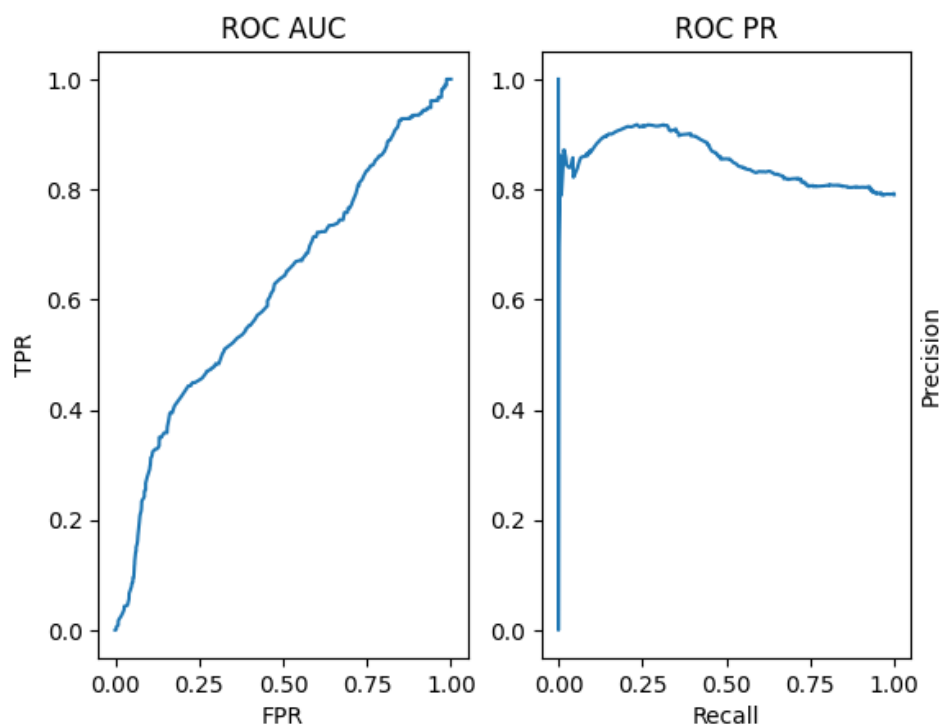
Fig. 40: Top 10 features with LOPO, binary classification: Catboost

Fig. 41: Top 10 features with LOPO, binary classification: LightGBM

Fig. 42: Top 10 features with LOPO, binary classification: SVM

Fig. 43: Top 10 features with LOPO, binary classification: Random Forest

Fig. 44: Top 10 features with LOPO, binary classification: MLP

Fig. 45: Top 10 features with K-folds, binary classification: Gradient Boosting

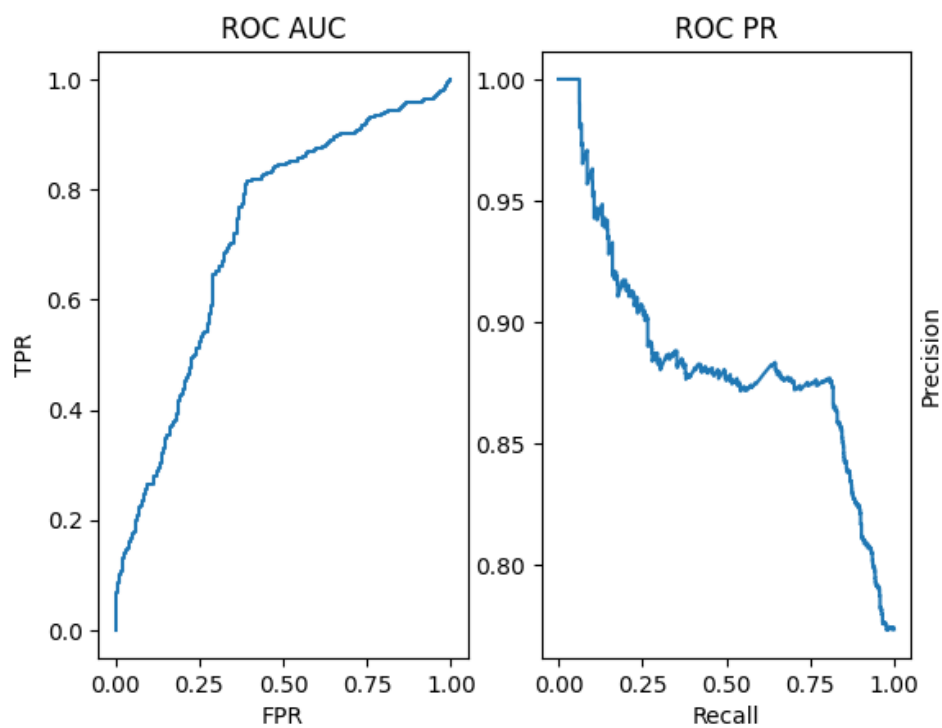Fig. 46: Top 10 features with K-folds, binary classification: Catboost

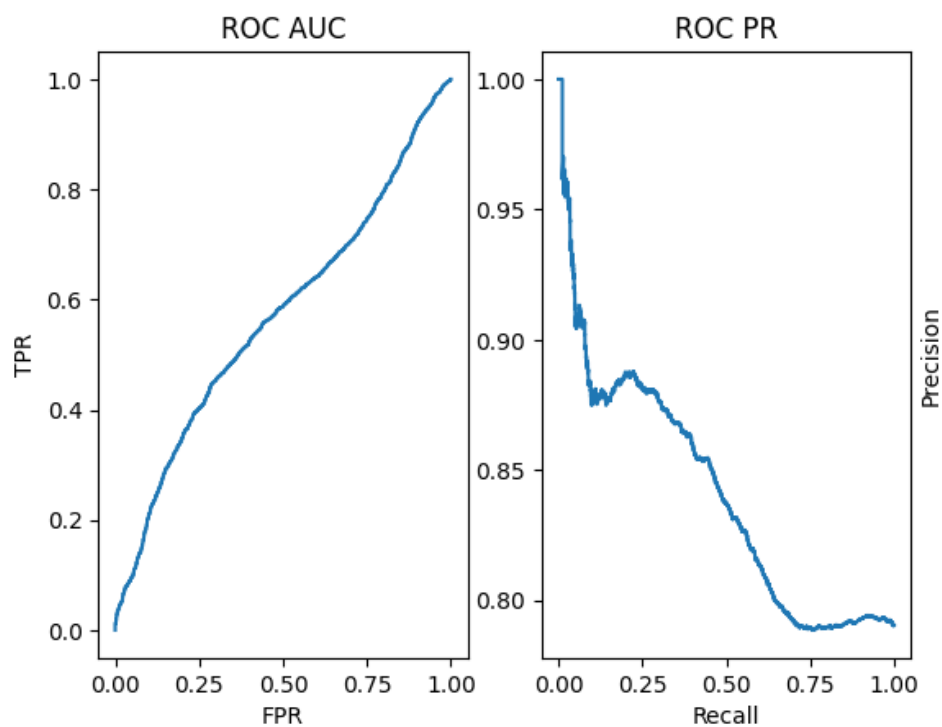Fig. 47: Top 10 features with K-folds, binary classification: LightGBM

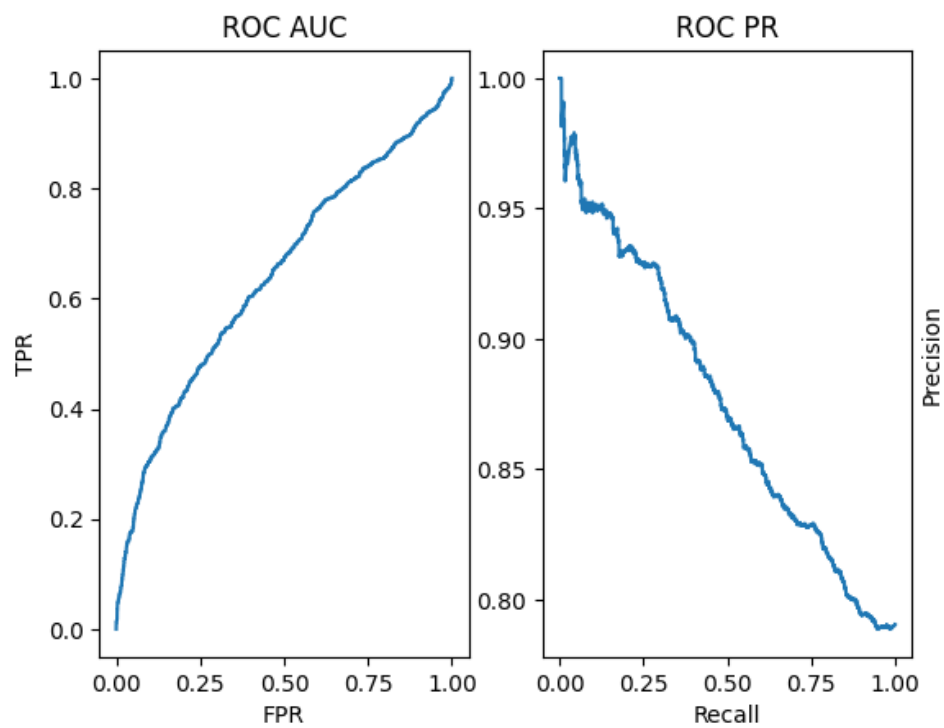Fig. 48: Top 10 features with K-folds, binary classification: SVM

Fig. 49: Top 10 features with K-folds, binary classification: Random Forest