Piano Mood Detect: Incorporating a Machine Learning Model with the Arduino BLE Sense to Analyze the Mood of Piano Music

DGMD S14
Harvard Summer School 2021
By: Alex Giannini, Kaif Jeelani, Isaac Song
08/03/2021

Table of Contents

1. Abstract 2
2. Introduction (adapted from proposal) 3
2.1. Data Collection 3
2.2. Visualization 5
2.3. Description 7
2.4. Related Work 8
3. Team Organization 9
3.1. Team Members and Roles 9
4. Software Developing Tools 10
4.1. Software 10
4.2. Laptop/Desktop setup 10
4.3. Hardware Needed 11
5. List of Milestones 12
6. Results and Discussion 14
7. Conclusions 17
8. Future Work 18
9. References 19

1. Abstract

After studying the Internet of Things (IOT) and Machine Learning (ML) in DGMD S14, the Piano Mood Detect (PMD) team has created an IOT device to classify audio using a ML model.

The ML solution leveraged the ‘Edge Impulse’ platform to create a Convolutional Neural Network (CNN) to execute a supervised learning multi-class classification algorithm on either streaming audio or short audio clips. To increase the accuracy, the model splits the audio into 1.5 second segments, extracts each clip’s Mel Spectrograph features and feeds the data to a neural-net-classifier powered by Keras. Then, the classifications are averaged over the 10 second time-period and the result is the clip’s mood (either happy, sad, or angry).

The hardware solution used Arduino’s BLE Sense. The chip was flashed using the Edge impulse CLI and the Arduino CLI and IDE. Then, the firmware code was modified to allow the device to process streaming audio and transmit the output to either a computer or a phone with BLE. As a result, the current prototype is mostly self-sufficient and only relies on the computer as a power source and a phone or computer for the display.

This report details the hardware and software considerations taken in the process of creating the PMD device and algorithm. It will describe the process of creating the final prototype and discuss how it proves the initial idea.

2. Introduction (adapted from proposal)

Franz Schubert stated “When I wished to sing of love, it turned to sorrow. And when I wished to sing of sorrow, it was transformed for me into love” (*BrainyQuote*). Even a musician as accomplished as Schubert recognized the mercurial nature of the artform. His declaration highlights how difficult it can be for musicians to communicate their intended emotion. This is because music is perceived according to a many unpredictable variables.

However, accurately predicting a person's emotional reaction to music is not impossible because sentiments are shared by large groups of people. Philosopher Émile Durkheim defined the collective consciousness as ‘the body of beliefs and sentiments common to the average of members of a society’ (*Oxford Reference*). This idea is a ‘fundamental sociological concept’ and highlights the ability to predict how a people will perceive anything (Cole). Therefore, if sentiments regarding music can be estimated, the reaction can be modeled.

2.1. Data Collection

To collect data for the training dataset, short 15-20 clips of piano music were recorded and uploaded to Edge Impulse using the console. The raw data can be found in the data acquisition tab of the edge impulse project and on the GitHub repository.

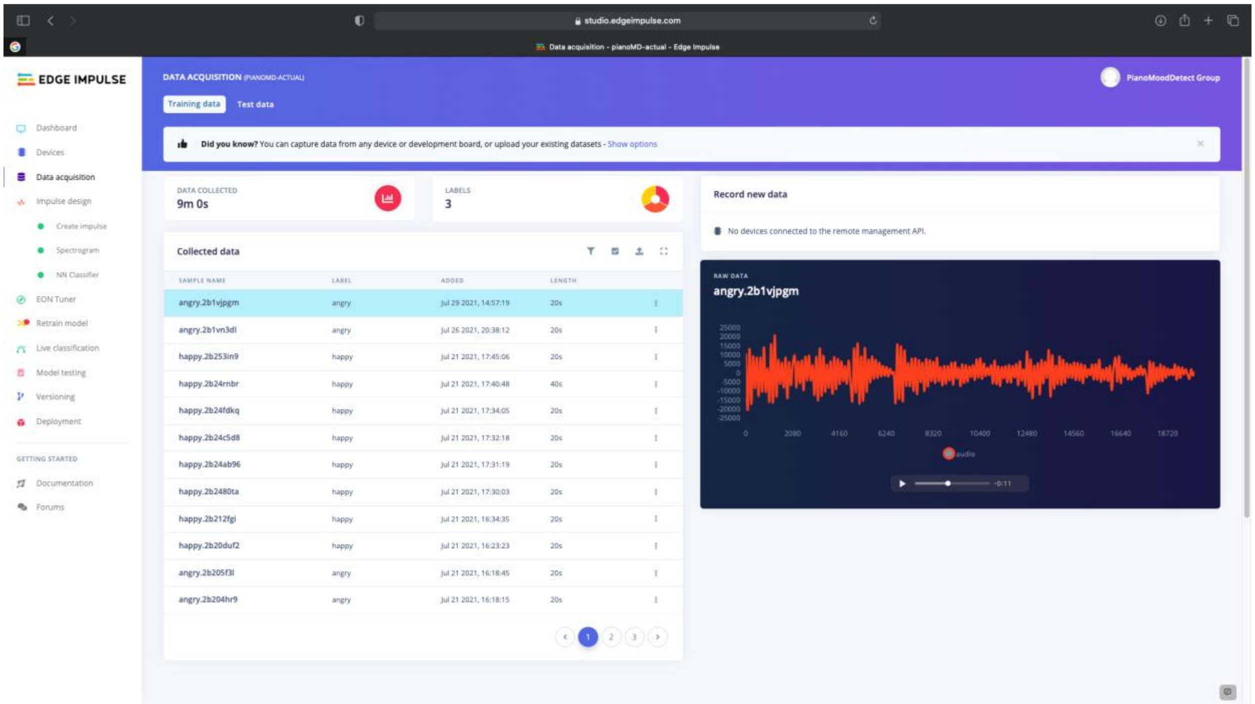


Figure 1: https://studio.edgeimpulse.com/studio/41648/acquisition/training?page=1

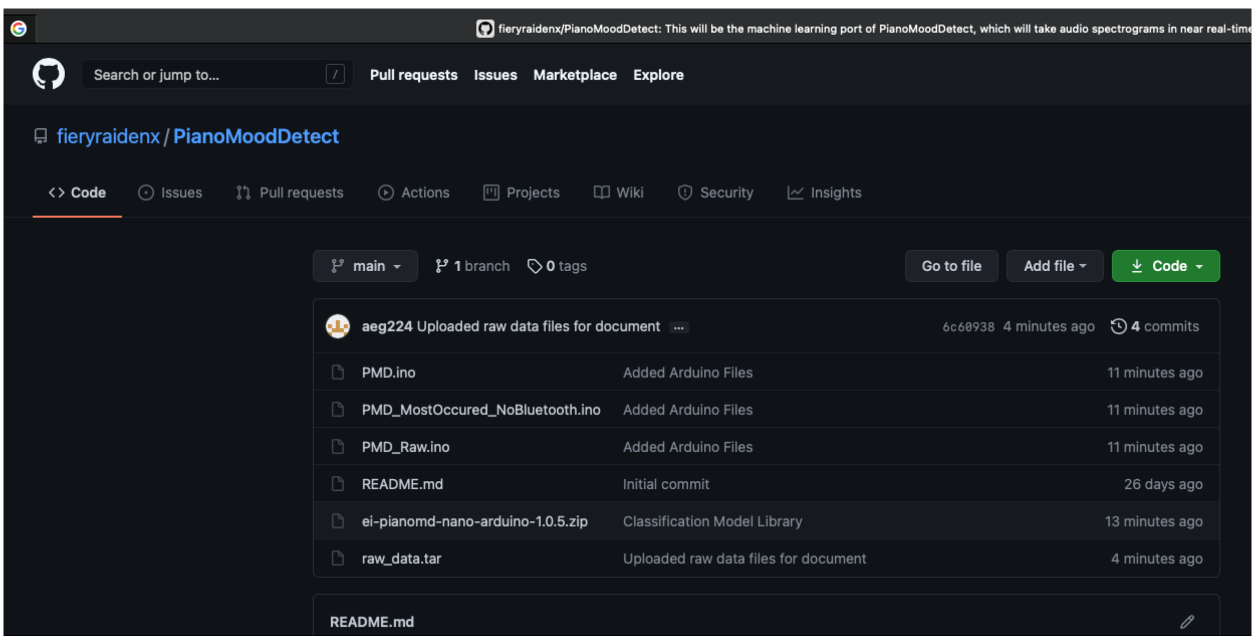


Figure 2: https://github.com/fieryraidenx/PianoMoodDetect

The training dataset consists of approximately 12 minutes of recorded audio, with three minutes and forty seconds dedicated to each mood and one minute of silence. The test dataset consists of a total of three minutes and twenty seconds of data. Three minutes are evenly split between the three moods and 20 seconds is dedicated to silence. Both datasets feature Kaif playing the piano and are recorded with an Arduino BLE Sense. Table 1 shows the breakdown of train data to test data.

Table 1: Train Data and Test Data

Type	Mood	Sensor	Total data	Musician
Train	Happy	BLE Sense	3:40	Kaif
Train	Sad	BLE Sense	3:40	Kaif
Train	Angry	BLE Sense	3:40	Kaif
Train	Silence	BLE Sense	40 seconds	Kaif
Test	Happy	BLE Sense	40 seconds	Kaif
Test	Sad	BLE Sense	40 seconds	Kaif
Test	Angry	BLE Sense	40 seconds	Kaif
Test	Silence	BLE Sense	20 seconds	Kaif

Initially, the train dataset was made of recordings created on the phone. But, to increase the accuracy, the recordings were re-made on the BLE-Sense. This decision will be discussed in more detail in the results section.

Also, the ratio of train to test data was carefully considered. At one point, there was a 90-10 percent ratio of train data to test data. However, after some research it was determined that the difference between the two should be lower and currently there is approximately 80-20 train to test ratio (Mitchell).

2.2. Visualization

Using Edge Impulse, the team was provided a visual representation of the model's execution of the test data which is shown in figure 3. The overall accuracy was measured at 87.24%. More details on the progression of the accuracy can be found on the 'results and discussion' section. Each datapoint on the graph represents a single 1.5 second clip. Red indicates an incorrect reading, while green indicates a correct one.

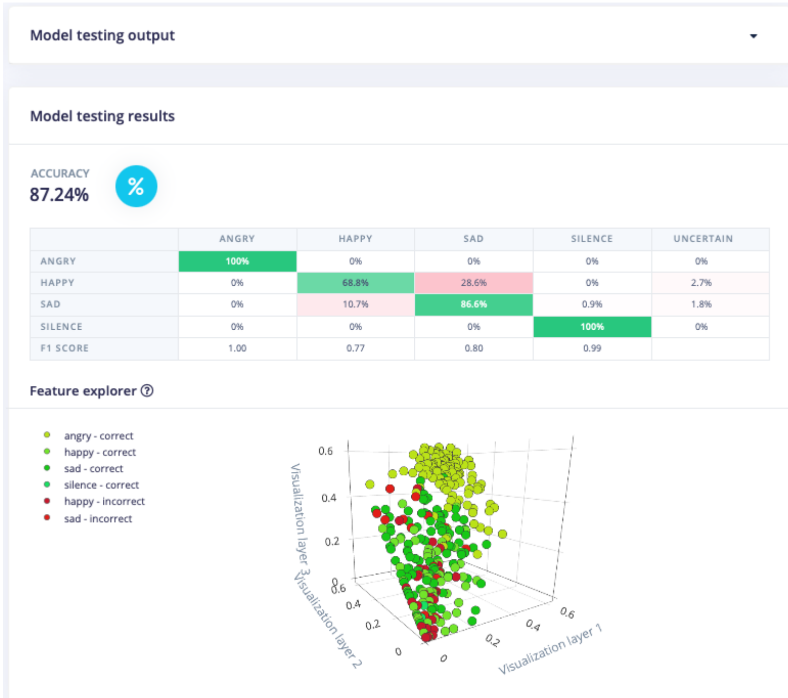


Figure 3: Test Data Visualization Provided by Edge Impulse

Because each clip of audio is analyzed at 1.5 second intervals, a single incorrect response does not indicate an incorrect conclusion. For example, the Figure 4 shows the output of a clip that was intended to be interpreted as ‘sad’. The algorithm took 56 readings, but the vast majority were determined to be sad clips so the overall conclusion was accurate.

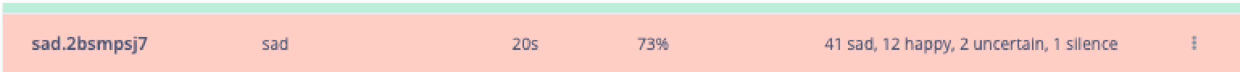


Figure 4: Each clip is analyzed into many datapoints

In addition, Edge Impulse has provided a visualization of the training data output, shown in Figure 5. The accuracy was measured at 97.8%. Each item in the chart represents an item of raw data.

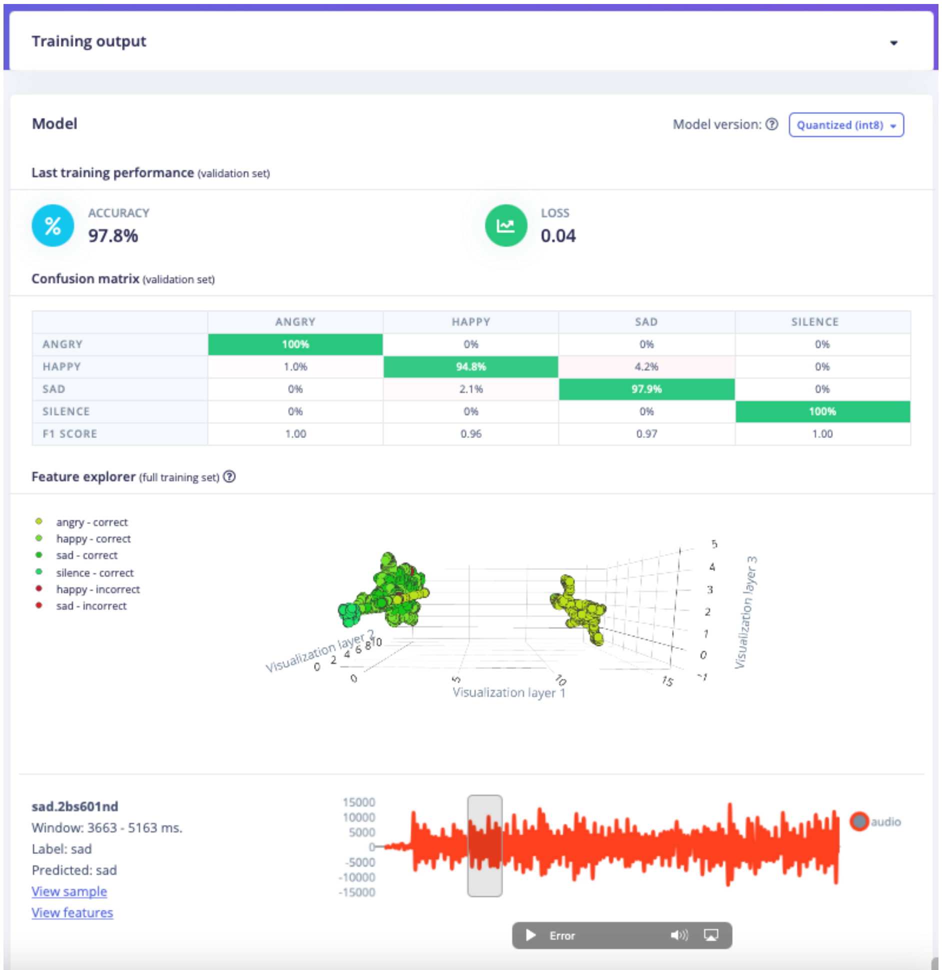


Figure 5: Train Data Visualization Provided by Edge Impulse

2.3. Description

The goal of project was to prove that MEMs sensors can be paired with a machine learning algorithm to create an IOT device that can be used identify the mood of the music it receives as input. Table 2, below, shows the initial feature list with an additional column added to indicate the final prototype's features.

Table 2: PMD Feature List

Feature	Secondary	MVP	MVP +	Long Term	Prototype
Audio Processing: Short Clips	Accurate Output with 15-20 second clip	YES	YES	YES	YES
Audio Processing: Long Clips	Accurate Output with clips > 30 seconds	NO	YES	YES	YES
Audio Processing: Streaming	Accurate output with streaming data	NO	NO	YES	YES
Music Classification: Happy	Classification: 80% or greater accuracy	YES	**YES**	YES	YES
Music Classification: Sad	Classification: 80% or greater accuracy	YES	YES	YES	YES
Music Classification: Angry	Classification: 80% or greater accuracy	NO	YES	YES	YES
Music Classification: Vibrant	Classification: 80% or greater accuracy	NO	YES	YES	NO
Music Classification: Calm	Classification: 80% or greater accuracy	NO	NO	YES	NO
Music Classification: Frantic	Classification: 80% or greater accuracy	NO	NO	YES	NO
Music Type: Piano	Accurate Output for Piano	YES	YES	YES	YES
Music Type: Other Strings	Accurate Output for Guitar, Violin, etc...	NO	YES	YES	NO

Music Type: Percussion	Accurate Output for Percussion	NO	YES	YES	NO
Music Type: Woodwinds	Accurate Output for Woodwinds	NO	NO	YES	NO
Music Type: Brass	Accurate Output for Brass	NO	NO	YES	NO
Music Type: Combinations	Accurate Output for Instrument Combos	NO	NO	YES	NO
Hardware: Audio Sensor	Will be used to collect data	YES	YES	YES	YES
Hardware: Algorithm Upload	Process data on the device	NO	YES	YES	YES
Hardware: Cloud Enabled	Send Data and Prediction to the Cloud	NO	YES	YES	NO
Hardware: Web Enabled	Send Prediction to Website	NO	YES	YES	NO
Hardware: Self Powered	Independent Power Supply	NO	NO	YES	NO
Embedded Hardware: Display	Attached Screen Display	NO	NO	YES	NO

All the MVP features were included in the final prototype along with most of the MVP +. In addition, the long-term goal of allowing for real time music processing by streaming the data was achieved. In the earliest planning phases, this was considered a main goal but eventually it was pushed to a long-term feature because it was determined to be beyond our capabilities. However, as the project grew, documentation detailing how to stream music was found and the feature was completed.

All the MVP+ features were achieved except for: Music Classification: Vibrant, Music Type: Percussion, Music Type: Guitar, Hardware: Cloud Enabled and Hardware: Web Enabled. The music classification and mood features were not included because they required a new musician to record new training and test datasets. Also, the Cloud and Web features were not completed so that streaming music, a long-term goal, could be focused on.

2.4. Related Work

During the creation of the prototype a variety of related work was discovered, including a paper titled ‘Music Mood Classification’ (Nuzzollo). The paper discusses its observations regarding techniques that can be used in music mood classification. While the PMD prototype uses an Edge Impulse model, the paper could have been particularly useful if the model had been built using Librosa because it references additional scholarly work on the subject. The chart below, which is from the Nuzzolo paper, details an approach for classifying music mood according to musical attributes.

Table 3: From Music Mood Classification by Nuzzolo : https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/

Table: **ood**	**Intensity**	**Timbre**	**Pitch**	**Rhythm**
Happy	Medium	Medium	Very High	Very High
Exuberant	High	Medium	High	High
Energetic	Very High	Medium	Medium	High
Frantic	High	Very High	Low	Very High
Anxious/Sad	Medium	Very Low	Very Low	Low
Depression	Low	Low	Low	Low
Calm	Very Low	Very Low	Medium	Very Low

Contentment	Low	Low	High	Low

In addition, a YouTube video titled “Extraordinary People - The Musical Genius (Full Show)” details a musician named Derek Paravincini who is an autistic and blind. The video highlights that people’s perception of music mood music vary greatly informs the discussion about the algorithm’s inherent bias (brainphreaky, 21:01). Also, the video demonstrates how perceptions regarding music can be measured by EEG Sensors which provided some insight in how inherent bias can be decreased (brainphreaky 18:45).

3. Team Organization

3.1. Team Members and Roles

In the earliest planning stages, it was decided that an Agile/Scrum approach would be best. According to Agile best-practice, the methodology was adapted to fit the unique requirements of the project, particularly the short amount of time, small team size and spontaneous meetings.

As a result, the roles of Product Owner, Developer and Scrum Master were assigned in addition to Subject Matter Expert (SME) roles. Tasks were assigned as a group, and the team coordinated via text or email if any impediments were found.

Kaif Jeelani

- Product Owner – Kaif had the initial idea and therefore was the best choice for Product Owner. He coordinated with the team to ensure the product fits with his vision.
- Developer - Kaif used his programming and machine learning skills to help develop a predictive neural network.
- SME: ML and Music – Kaif’s previous experience with Machine learning made him the team ML SME. He was also team’s pianist and recorded the test and training data.

Alex Giannini

- Scrum Master – Alex’s previous experience with Agile made him the best choice for Scrum Master. Alex helped facilitate an agile approach to the project and created and managed the Jira board.
- Developer - Alex used his programming skills and knowledge of computer science to help develop a predictive neural network.
- SME: Writing and Agile – Alex’s experience with Agile and writing research reports made him the best choice for Writing and Agile SME. He was responsible for proofreading and writing the team documents.

Isaac Song

- Developer - Isaac used his machine learning skills to help develop the predictive neural network.
- SME: Arduino BLE Sense – Isaac had the most experience creating hardware with Arduino boards and therefore was the best choice for hardware specialist.
- SME: Arduino Coding – Isaac used his Arduino coding experience to incorporate the audio streaming features.

4. Software Developing Tools

A wide range of tools were used to develop the prototype. As the prototype evolved, some of the software changed. This section details the required software as well as the software used along the way.

4.1. Software

Building the prototype requires the following software.

Python 3
Node.js v14
Edge Impulse CLI
Edge Impulse Console
Home Brew (Mac OS/Linux)
Arduino CLI

The official software download for each is listed below. This software was chosen after experimenting with the ST Electronics Software and IDE. While this informed our decisions, it was not used in the prototype.

https://www.python.org/downloads/
https://nodejs.org/en/
https://docs.edgeimpulse.com/docs/cli-installation
https://www.edgeimpulse.com
https://brew.sh
https://arduino.github.io/arduino-cli/latest/installation/
https://www.arduino.cc/en/software

Initially, the System Workbench for STM32 and Blue Coin Firmware Packages were considered. However, for reasons discussed in the Hardware section, the project switched from using an ST Electronics Solution to an Arduino solution and the software list was modified to the list above. While they were not necessary to build the final prototype, they helped determine the best option moving forward. Their links are also included below.

https://www.st.com/content/st_com/en/products/embedded-software/mcu-mpu-embedded-software/stm32-embedded-software/stm32-ode-function-pack-sw/fp-aud-bvlink1.html
https://www.st.com/en/development-tools/sw4stm32.html

4.2. Laptop/Desktop setup

The prototype can be created with either a Mac or a PC. Only 1 free USB port and a USB to Micro-USB Cable is required. They are used to display the output, provide power, and upload the model to the hardware.

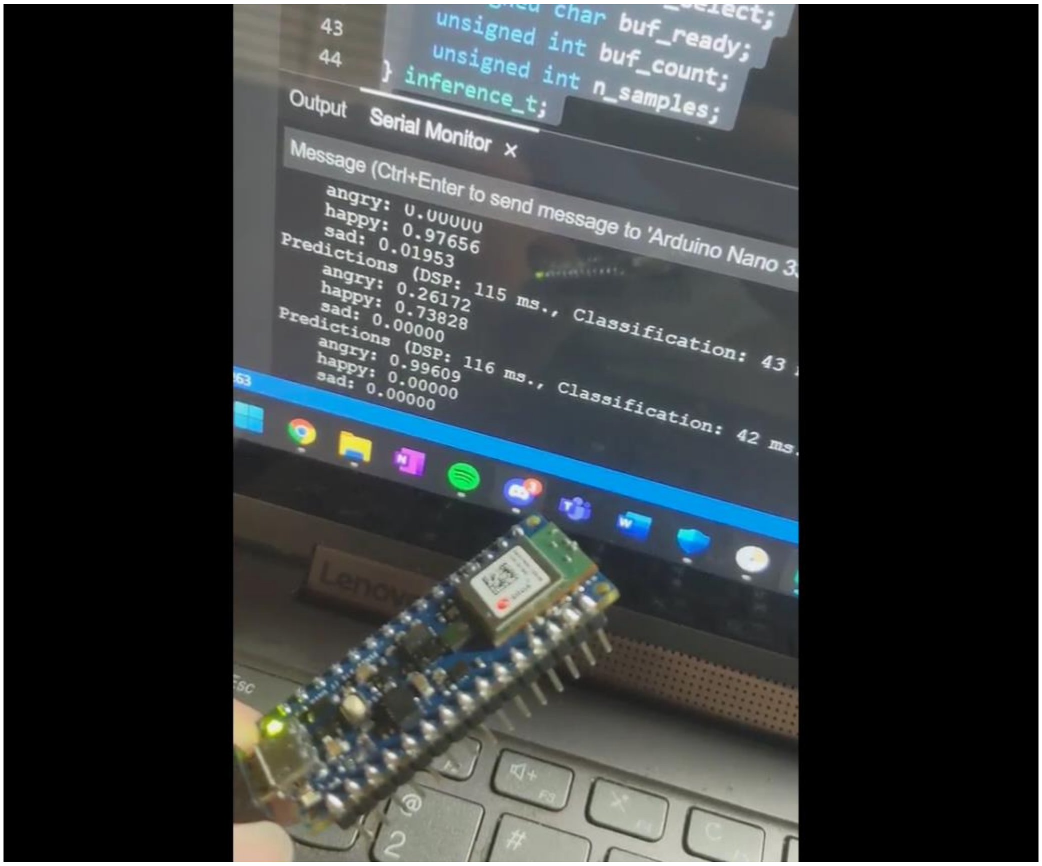


Figure 6: Picture of prototype hardware

Once the prototype is created and configured, the computer is only used to provide power and a display. If a power supply and display were attached to the hardware, a computer would only be necessary to configure the device.

In addition, a phone can optionally be used as a display. The Arduino transmits the output via BLE to an Arduino application on the phone. If the phone is used the computer is only used to provide power and could also be replaced with a power supply.

4.3. Hardware Needed

The prototype uses the Arduino BLE Sense (pictured below).

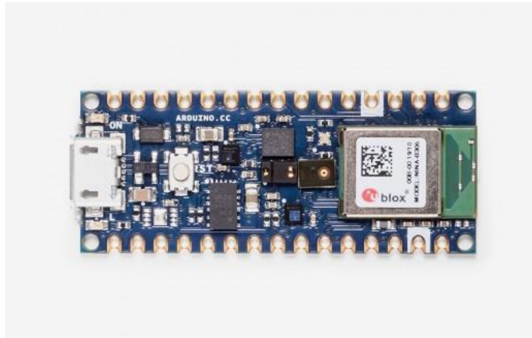


Figure 7: From the Arduino Webside - https://store.arduino.cc/usa/nano-33-ble-sense

Initially, the ST Electronics Sensor Tile and Nucleo Board were considered. However, due to the microphone quality and the complexity of the firmware, we began to explore other options. The ST Electronics Blue Coin with Nucleo Board was one option, but after analyzing the firmware, it was also ruled out. Shortly after, the Arduino BLE Sense was chosen due to its low-cost, physical size, ample memory, high-quality sensors and support for the Edge Impulse CLI.

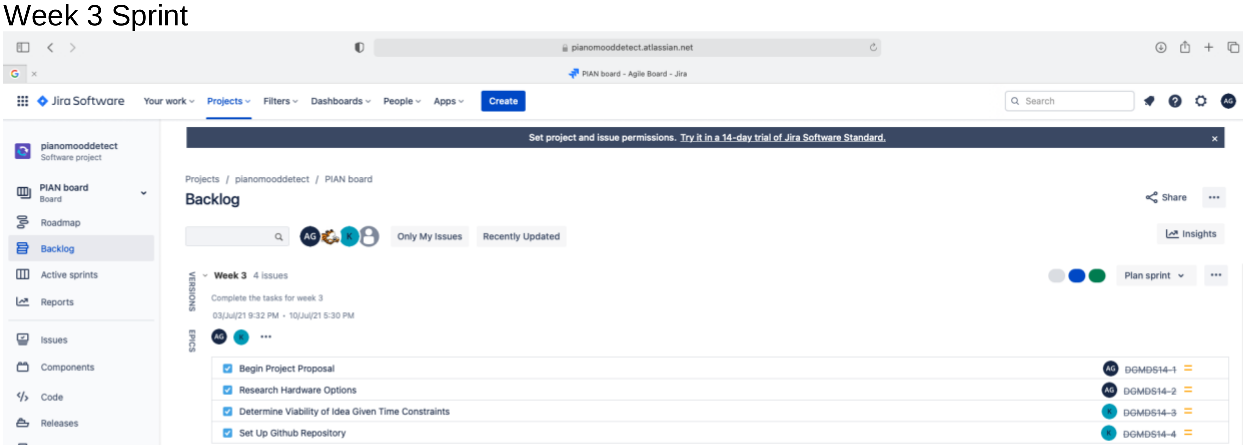
While the prototype does not utilize the SensorTile, Blue Coin or Nucleo Board, they were an important part of the decision-making process. More information on the ST Electronics Blue Coin can be found on the st website (link below).

https://www.st.com/en/evaluation-tools/steval-bcnkt01v1.html

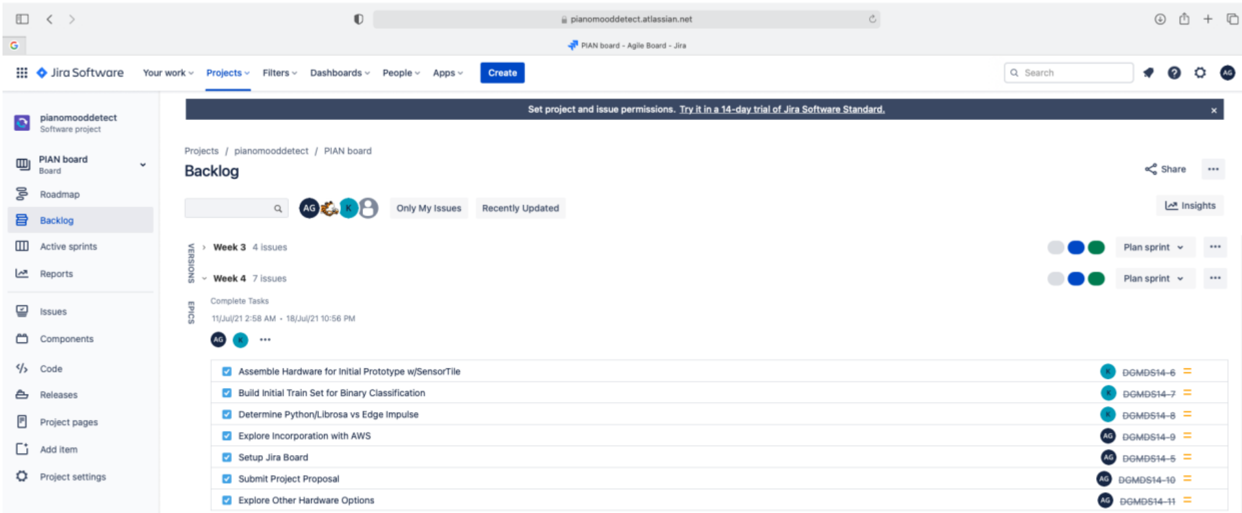
5. List of Milestones

The team attempted to take an Agile approach to task management. Each week was considered 1 sprint. The tasks were discussed and assigned as a group which allowed each team member to work in areas of interest and knowledge. Once the tasks and assignments were decided, they were uploaded to Jira. The Jira board can be found at https://pianomooddetect.atlassian.net/secure/RapidBoard.jspa?rapidView=1&projectKey=DGMDS14&view=planning.nodetail&selectedIssue=DGMDS14-32&issueLimit=100

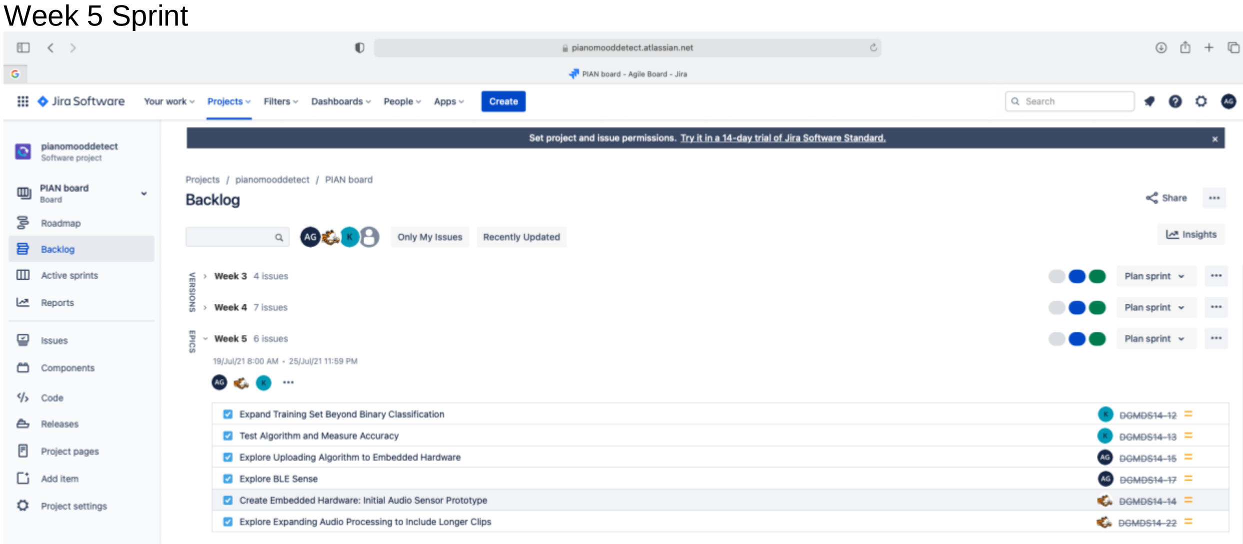
Week 3 Sprint



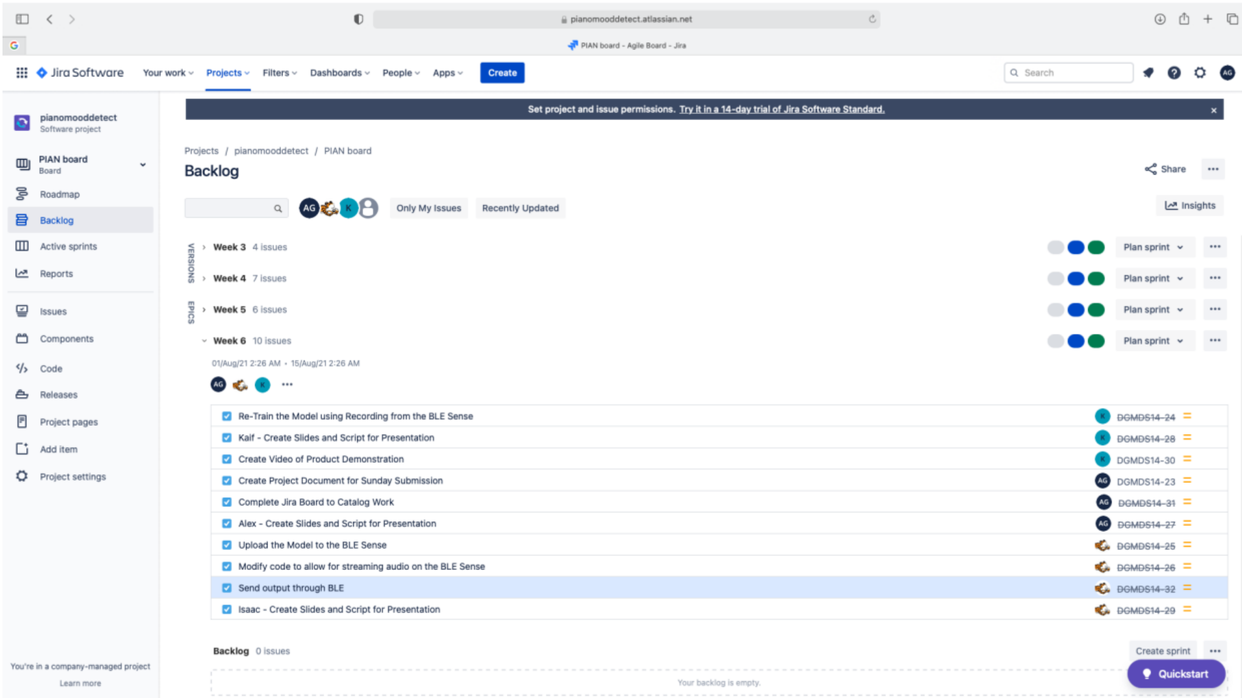
Week 4 Sprint



Week 5 Sprint



Week 6 Sprint



6. Results and Discussion

The final prototype contains all the features listed in Table 4. These include all the MVP and most of the MVP+ features detailed in Section 2.3.

Table 4: ProType Feature List

Feature	Description
Audio Processing: Short Clips	Accurate Output with 15-20 second clip
Audio Processing: Long Clips	Provides Accurate output clips > 30 seconds
Algorithm Processing: Streaming	Provides Accurate output with streaming data
Music Classification: Happy	Classification: 80% or greater accuracy
Music Classification: Sad	Classification: 80% or greater accuracy
Music Classification: Angry	Classification: 80% or greater accuracy
Music Type: Piano	Accurate Output for Piano
Hardware: Audio Sensor	Will be used to collect data
Hardware: Algorithm Upload	Allows device to process data on the device

The Edge Impulse Model first structured the received data in 1.5 second bursts. Then, the mel-spectographic features are extracted and fed into a 1D-CNN classifier created by Keras. Finally, it is classified into one of four outputs: angry, happy, sad and silence. This sequence is detailed in Figure 8 below.

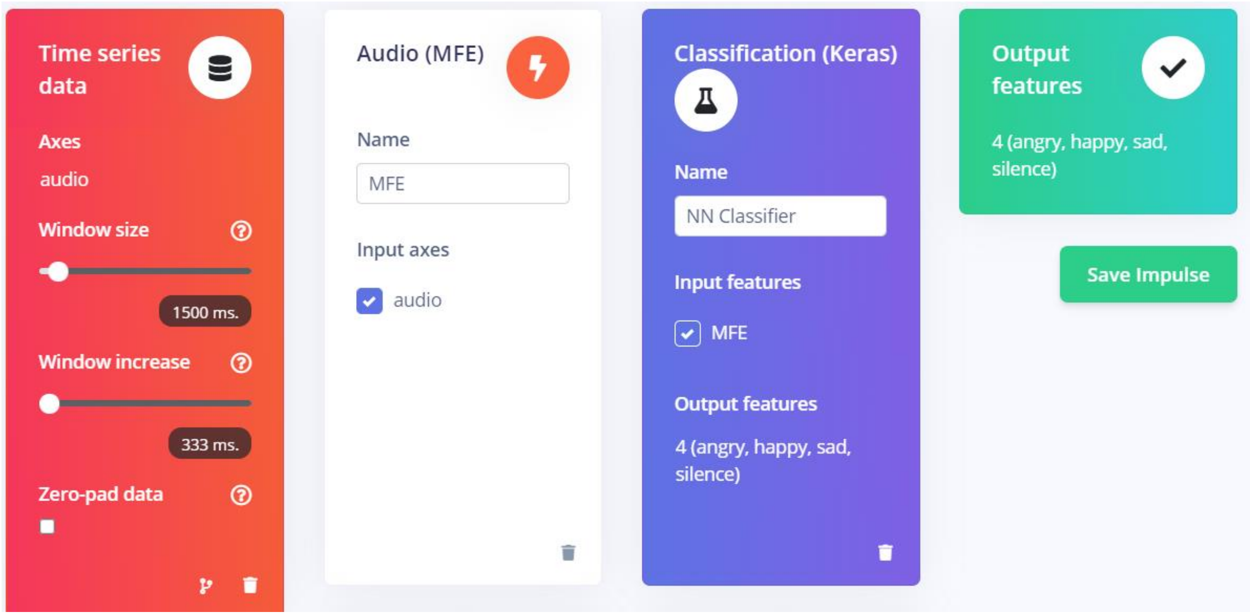


Figure 8: Edge Impulse Models Sequence of Processing

The Keras classifier has: an input layer to receive input, a reshape layer to process the data, a 1D Conv/Pooling layer which allows features to be extracted and processed, a dropout layer which prevents overfitting by randomly dropping neurons in training. These are shown in Figure 9 below.

Figure 9: Keras Classifier Layers

A dataset was first built to include classification for four moods; however, this produced a test accuracy of around 45%. After meeting to discuss the results, the team decided to modify the scope of the project to only include support for three moods. This produced an increased accuracy of about 25% resulting an overall test accuracy of 70%.

Next, the model was modified to manage silence. Most of the recordings in the dataset contained some silence but the train dataset for the mood ‘sad’ contained the most silence out of the three. As a result, it was classifying all silence as ‘sad’ and clips with no sound as sad. To solve this problem, a silence tag was created to devalue silence in the algorithmic calculations and prevent a silent clip from being labeled as a mood. This can be seen in Figure 10.

Essentially, the model treats silence as its own mood and therefore does not influence the average. The result of this change was an increase an additional 10% to achieve an 80% test accuracy.

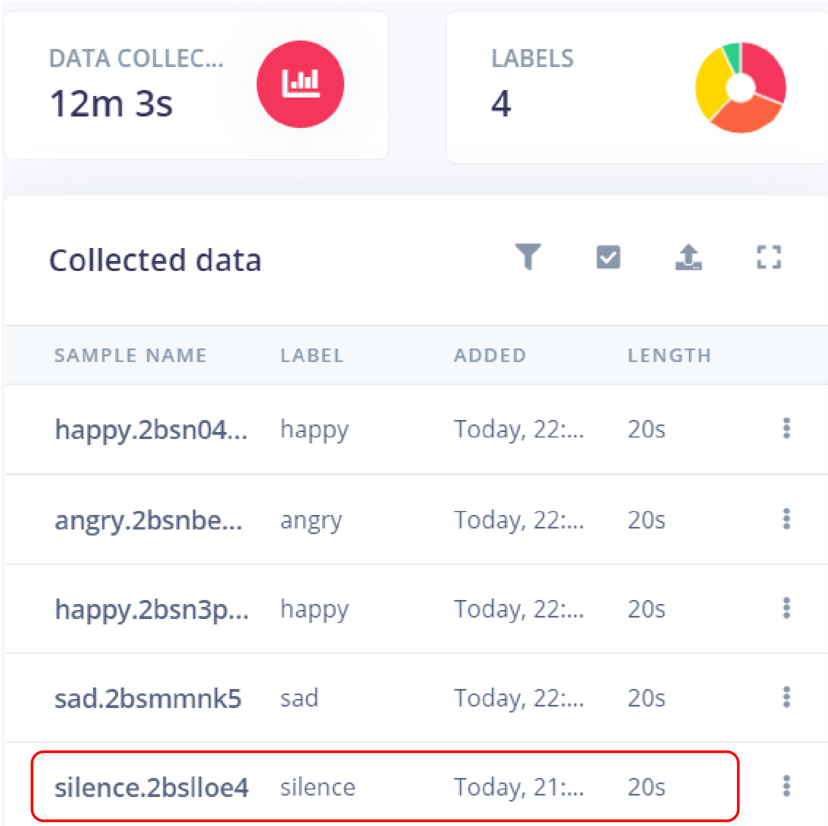
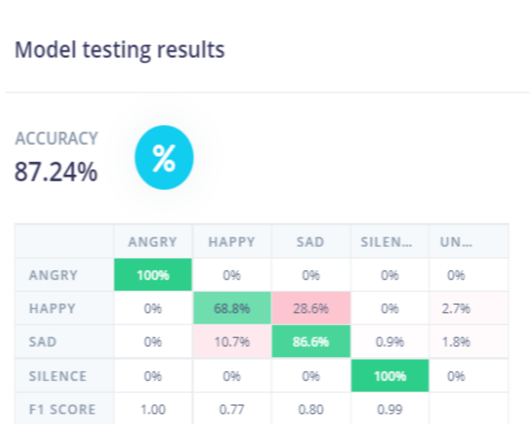


Figure 10: Silence Added to Edge Impulse Model

Then, to improve the model further, the dataset was re-recorded, and the processing was modified. Initially, the model was using the Spectrographic data. However, when the model was retrained using MELs information, it slightly increased.

Also, the initial dataset was created by using the phone to record the data. However, when streaming was incorporated with the Arduino BLE Sense, it was clear that all test data outside of the test dataset was to be recorded by the Arduino microphone. As a result, the whole data set was re-recorded producing an accuracy of 87.24%. Figures 11 and 12 show the Edge Impulse models accuracy results for both test and training.

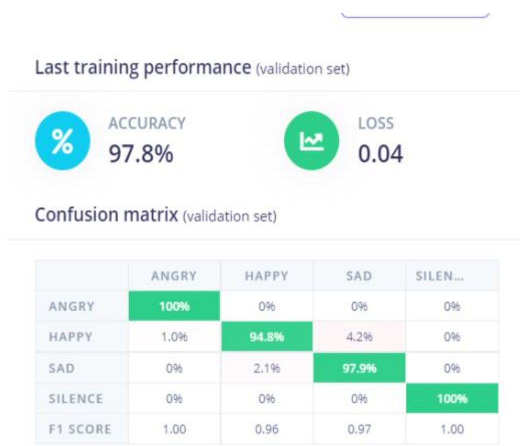


Model testing results

ACCURACY
87.24%

	ANGRY	HAPPY	SAD	SILEN...	UN...
ANGRY	100%	0%	0%	0%	0%
HAPPY	0%	68.8%	28.6%	0%	2.7%
SAD	0%	10.7%	86.6%	0.9%	1.8%
SILENCE	0%	0%	0%	100%	0%
F1 SCORE	1.00	0.77	0.80	0.99	

Figure 11: Edge Impulse Model Testing Results



Last training performance (validation set)

ACCURACY
97.8%

LOSS
0.04

Confusion matrix (validation set)

	ANGRY	HAPPY	SAD	SILEN...
ANGRY	100%	0%	0%	0%
HAPPY	1.0%	94.8%	4.2%	0%
SAD	0%	2.1%	97.9%	0%
SILENCE	0%	0%	0%	100%
F1 SCORE	1.00	0.96	0.97	1.00

Figure 12: Edge Impulse Model Training Results

7. Conclusions

After reflecting on the project, the PDM team feels confident the prototype meets the goals set in the project's outset. The initial idea was to prove that a music mood classifier could be created with streaming audio using a small board with MEMs sensors. And, while the project only deals with piano music and three moods, the same Machine Learning principles could be applied to expand the functionality to other moods and instruments.

Upon further reflection, the team feels the process of creating the prototype has greatly enhanced everyone's knowledge in the areas of ML, MEMs hardware, audio processing and merging hardware with a ML model.

The project has enabled the PMD team to enhance their knowledge of ML. To determine the best course of action, a variety of ML approaches were researched. These include DSP with python and Librosa, AWS based architectures with Sage Maker and S3, and different approaches for the Convolutional Neural Network. While these were not all used, the research widened the breadth of understanding on the general subject.

In addition, the project has enabled the PMD team to enhance their knowledge of MEMs hardware. There were several meetings regarding chip selection, and many were considered. Since the class primarily uses ST Electronics devices, settling on Arduino came with some risk. However, since one of the teammates had some previous experience, it was determined to be worth the risk and resulted in the whole team gaining practical experience with an additional chip set.

Furthermore, the PMD team gained experience merging hardware with a ML model. While this process was greatly facilitated by Edge Impulse, the act of flashing a chip with an ML model forced the team to think critically about embedded systems and their limitations. While the decision to upload the model on the chip was relevant for the use-case there was some discussion about alternative options especially if the ML model were intellectual property (IP) or too large for the chip set.

8. Future Work

The PMD prototype was designed to work within the time-restrictions of DGMD S14. However, if a longer period were available, the team would have included support for a more music and moods, leveraged the cloud and limited our bias by conducting a separate research project.

To expand the range of music and moods a new data set would have to be created with substantial data for each mood. A 2017 study from UC Berkley's Greater Good Science Center determined there were as 27 distinct human moods (Keltner). If each of these moods translated to music, then data for 27 moods would be required multiplied by each instrument type (strings, woodwinds, keys, brass, percussion) plus the combinations of instruments. This would mean well over 200 different categories with each requiring a significant amount of audio recording. In addition, to make a robust model, the recordings should be played by a wide range of musicians in each category.

Also, the model would have been re-architected with the cloud and Librosa. Using a library like Librosa would allow more granular control over the model. For example, if it were determined that a regular spectrogram would give more accurate reading than a MEL spectrogram on a specific category, Librosa would allow the change for just the one category, while Edge Impulse would modify the whole model. Also, with a Librosa based model the processing could be more easily moved to the cloud. The model could be uploaded to AWS Sage maker and the sensor could be configured to stream audio to an AWS S3 bucket, which would then pass the data to the model and send the response to mobile app using AWS Edge Compute. This new architecture would protect the model's IP and would allow the model to expand without the limitation of the chip size.

Furthermore, future iterations could address the model's inherent bias. The algorithm uses the team's personal suppositions regarding the mood a piece. This is relevant because emotion is subjective as indicated in a recent Psychology Today article that declared "everyone has different types of emotional truth" (*Psychology Today*).

Therefore, it would be essential to conduct a separate study measuring a large and diverse group's reaction to music. To ensure the integrity of the subject's emotional response, biometric solutions could be used. Several companies offer biometrics including iMotions.com, which provides real-time eye tracking measurements, facial expression analysis and Electrodermal activity. In addition, an EEG could be used to measure brainwave responses. While most EEG's are for medical use, openbci.com and neurosky.com both provide customizable EEG options.

To execute the study, volunteers would need to wear a combination of these biometric devices while listening to music from a pre-determined list of categories. To ensure the data does not reproduce their bias, there would need to be many volunteers from a diverse background. Once the data is collected, it could be used to build a model that will determine the appropriate classification of each music without relying on our own suppositions.

9. References

"Franz Schubert Quotes." *BrainyQuote*, Xplore, www.brainyquote.com/quotes/franz_schubert_175104.

“Collective Conscience.” *Oxford Reference*, www.oxfordreference.com/view/10.1093/oi/authority.20110803095623983.

Nicki Lisa Cole, Ph.D. “Understanding Emile Durkheim's ‘Collective Consciousness.’” *ThoughtCo*, www.thoughtco.com/collective-consciousness-definition-3026118.

Mitchell, Tom M. “Machine Learning.” *Amazon*, McGraw Hill, 2017, docs.aws.amazon.com/machine-learning/latest/dg/splitting-the-data-into-training-and-evaluation-data.html.

Nuzzollo, Michael. “Music Mood Classification.” *Electrical and Computer Engineering Design Handbook*, sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/.

brainphreaky. “Extraordinary People - the Musical Genius (Full Show).” *YouTube*, YouTube, 10 Jan. 2012, www.youtube.com/watch?v=fibZudrZUto&t=1161s.

Keltner, —Dacher, et al. “How Many Different Human Emotions Are There?” *Greater Good,* greatergood.berkeley.edu/article/item/how_many_different_human_emotions_are_there.

“How to Know Your Emotional Truth.” *Psychology Today*, Sussex Publishers, www.psychologytoday.com/us/blog/the-empowerment-diary/201904/how-know-your-emotional-truth.

Figures 3, 4, 5, 8, 9, 10, 11 and 12 are pictures from the Edge Impulse Model. This can be seen at https://studio.edgeimpulse.com/studio/43615/ . If needed, contact Kaif, Alex or Isaac for login information.