

# 音解析による唐揚げの揚げ具合判定手法

陳 維<sup>1</sup> 塚田 紘也<sup>1</sup> 横窪 安奈<sup>1</sup> ロペズ ギヨーム<sup>1</sup>

**概要:** 自宅で料理する機会はコロナ禍に増加していたことから, 火加減の調節や調理の提示など現代社会の料理に対する支援の必要性が増えることが分かった. 従来研究として画像, 加速度センサ, 温度センサなどを利用して料理行動を理解することや調理中の音を解析して調理状態を判断する研究があるが, 火加減の調節や調理手順提示に関する多くの料理支援研究はセンサを調理器具につけているため, 料理中に水がかかけられる, 設置が不自由など問題がある. 本研究はスマートフォンの内蔵マイクから音響信号を取得し, 特に中身が熟しているかが分かりづらい唐揚げの調理音を選定し, 唐揚げの上がり具合を判断する料理支援を目的とする. 揚げ具合を揚げはじめ・変化 1・変化 2・変化 3・揚げ上がり 5つのラベルと定義する. 音響信号の特徴量として MFCC を抽出し, TPOT による機械学習分類した結果全体の F1-score が 97.14%, 揚げ上がりの F1-score が 85.71%まで達し, 音響的揚げ物の揚げ具合推定の有効性を示した.

## Fried Chicken Fried Condition Detection by Acoustic Analysis

CHEN WEI<sup>1</sup> HIROYA TSUKADA<sup>1</sup> ANNA YOKOKUBO<sup>1</sup> GUILLAUME LOPEZ<sup>1</sup>

### 1. はじめに

コロナ禍の在宅時間の増加をきっかけに現代社会の料理に対する向き合い方も変化している. 2020 年農林水産省による「食育に関する意識調査」[1]から, 約4分の1の人が自宅で料理を作る回数が増えたと答えた, 特に若い世代において割合が40%まで上る. 料理需要の増加により初心者向けの料理支援の必要性もあると考えられる. 料理初心者の大きな壁として曖昧さや技術などが挙げられる. 曖昧さに関しては調味料の量に使用される少々や適量・火の加減における食材の状態変化の見極めが, 技術に関しては魚の三枚おろしをはじめとした多様な包丁さばきなどが料理初心者にとっての大きな壁となっている. 料理初心者のためにすぐに習得することができない技術や判断を自動的に行えるスマートキッチンが料理する負担を減らし, 料理する過程を楽しませることができる. また近年 IoT(Internet of things)の発展はスマートフォンやタブレットだけでなく日常のあらゆるものにセンサが搭載されている. 調理器具, キッチン用品などに IoT を導入するケースも少なからず存在する. [2]では, 調味料の容器にモーションセンサ,

LED ライトを用いることで調味料追加の支援システムを提案した. [3]では, 包丁の柄先に加速度センサを用いてレシピ動画の自動同期システム SynCook を提案した. 料理と IoT の融合により, 日々の料理を可観測, 可制御にすることができるようになっている.

本研究は揚げ物の料理する際の出来具合を着目している. 揚げ物は食品に衣とパン粉をコーティングし, 高温の油に浸され表面の水分を蒸発させることでパリパリの食感を実現している. 周りの衣とパン粉のサイズ, 揚げる物体のサイズ, 揚げる時間によってテクスチャーのクリスピー度も変わる [4]. そのため表面の衣が揚げたに対して中身が揚げ切っていないケースもあり, 表面だけで揚げ物の揚げ具合を判断することが難しいと考える. 揚げ物を揚げる際, 衣の水分損失が指数的に減少すること [5]から, 水分と油の振動によって揚げ物の揚げ具合に音響的变化があると想定される. また, スマートフォンの内蔵マイクなど音響的センサは身近な機械に組み込むことが容易であり, 料理中に水がかかけられるやセンサの設置が不自由などの問題も解消される.

本研究はスマートフォンの内蔵マイクから音響信号を取得し, 特に中身が熟しているかが分かりづらい唐揚げの調理音を選定し, 唐揚げの上がり具合を判断する料理

<sup>1</sup> 青山学院大学  
Aoyama Gakuin University

支援を目的とする。背景雑音をフィルタリングして除去し、音響的特徴量として MFCC(Mel-Frequency Cepstral Coefficient) を抽出した。最後に機械学習を利用して唐揚げの揚げ具合の分類を行いモデルの評価をした。料理支援と音響的技術を組み合わせ、料理の曖昧さを改善する手法を提案し、料理初心者でも唐揚げを美味しく揚げるができる手法の考察を行った。

## 2. 関連研究

Yusaku らは音響信号に基づく料理状態の認識について検討していた [6]。実際の料理過程から収集されたデータセットを動作から切る、炒める、その他の 3 つのクラスに分け、料理状態の推定を行った。音響的特徴量として MFCC と NMF(Non-negative Matrix Factorization) を GMM(Gaussian mixture models) と混合したモデルを使用した。MFCC を用いたモデルではフレームごとの F1 値の精度が 70% を達した。料理行動認識の課題は特定の料理に限る場合は多いため、課題の多くは汎用する動作の認識である。本研究は目的は揚げ具合の推定のため、分類ラベルの定義などの再検討が必要とする。

## 3. 実験用データ

本節では行ったデータ取得とその評価について述べる。



図 1: 各段階の唐揚げ変化ラベル

### 3.1 データ収集

唐揚げを揚げている様子を図 1 のように録画データを取得した。データの取得は揚げ具合解析と揚げ過ぎ解析の 2 種類を行った。揚げ具合解析は音データの取得のために、揚げ過ぎ解析は揚げ過ぎることによる変化と揚げ物の揚げ上がりの条件の推定のために行った。実験環境は静かなキッチン環境内で行い、鍋の上から撮影した。録音に使用した機器は iPhone11 の一般的な録画機能で、温度管理のために揚げもの用温度計クックサーモ 5495B クロム (TANITA 製) を使用した、音声のサンプリング周波数は 16000Hz に設定した。唐揚げを温度別、個数別、油使用回数に分け計 10 回揚げた。10 回分の計測の詳細は以下の表 1 に示す。データを取得する際に 170 °C を一般的な平均温度に、7 個を普通の量と定義したうえで量多め (10 個)、量少なめ (5 個)、低温、高温を網羅するようにした。また、

油の使用回数によって変化があるかどうかを確かめるために同じ条件で 2 セットずつ揚げた。唐揚げの揚げ上がりの判断は経験ありの料理人の判定で行った。

表 1: 計測 10 回分の唐揚げの揚げ方

	温度 (°C)	個数	油回数
動画 1	170	7	1
動画 2	170	7	2
動画 3	170	10	1
動画 4	170	10	2
動画 5	170	5	1
動画 6	170	5	2
動画 7	150	7	1
動画 8	150	7	2
動画 9	200	7	1
動画 10	200	7	2
動画 11(揚げすぎ)	170	2	1

### 3.2 揚げ具合ラベル

図 1 の揚げ具合解析のラベルは以下の条件によって分類する。取得したデータは音声解析ソフト ELAN を使用してラベリングを行った。揚げ具合解析のラベルは揚げはじめ・変化 1・変化 2・変化 3・揚げ上がりの 5 つのラベルを取得し、窓サイズは 10 秒とし、シフトサイズ 1 秒のオーバーラップを行った。

- 「揚げはじめ」では完全に沈んだ状態で気泡のほうが目立つ状態。
- 「変化 1」では沈んだ状態で色に変化なしが、音調と音量の変化を明確に感じる。
- 「変化 2」では沈んだ状態で気泡のほうが目立つ、少しきつね色が付き始める。
- 「変化 3」でははっきりきつね色が付き、浮く始めるようになる。
- 「揚げ上がり」ではしっかりとした色が付き、完全に浮いている状態となる。

### 3.3 揚げ具合解析

まずすべての動画で同じ傾向が見られたために代表として一般的な条件である動画 1 (温度 170 °C、個数 7 個、油 1 回目) を使用してラベルごとの変化を示す。フーリエ変換による変化の様子は図 2 のようになる。この図は左上からラベル (揚げはじめ、変化 1、変化 2、変化 3、揚げ上がり) ごとの変化の様子を表している。ラベルごとに振幅スペクトルの量が増えていることがわかる。さらに高周波数帯の振幅スペクトルの増加も見られる。

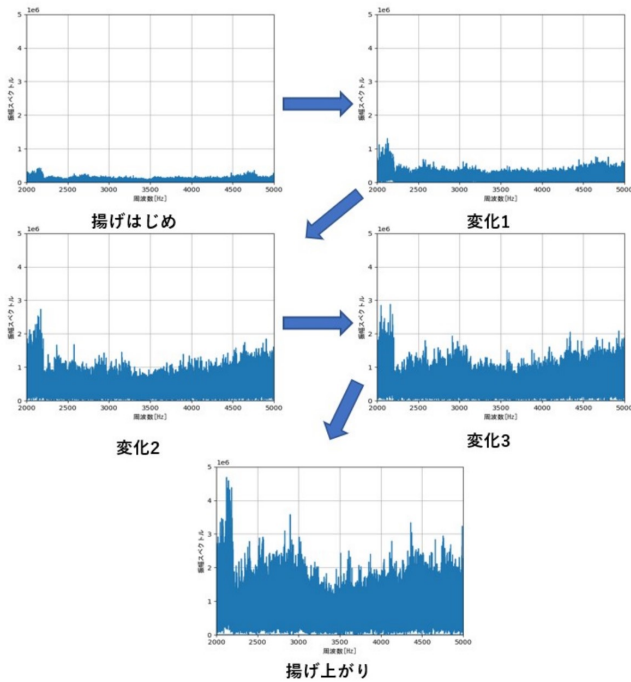


図 2: フーリエ変換による変化の様子

スペクトログラムとメルスペクトログラムによる変化の様子は図 3a と図 3b のようになる。黒線は時間ラベルの境界線に合わせている。それぞれの図のワースペクトルに注目すると、全体的な周波数と徐々に高い周波数帯両方が増加しているため、揚げ上がりのラベルに近づくにつれて色が濃くなっていくことがわかる。

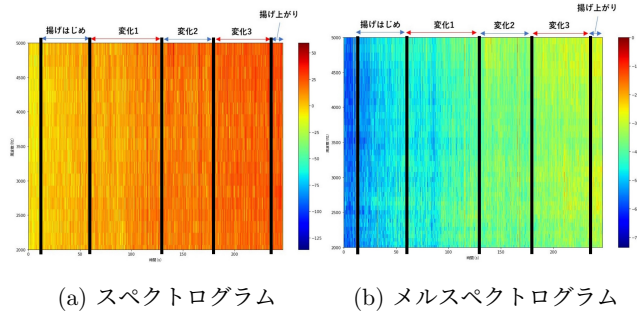


図 3: 時間別周波数帯域

さらに図 4 から平均振幅スペクトルの増加を確認することができた。揚げはじめから揚げ上がりにかけての平均振幅スペクトルは単調増加していくが、油の温度、唐揚げの個数が変わると平均振幅スペクトルの基準も同時に変化するため、揚げ上がりの判定の基準として満たさない。

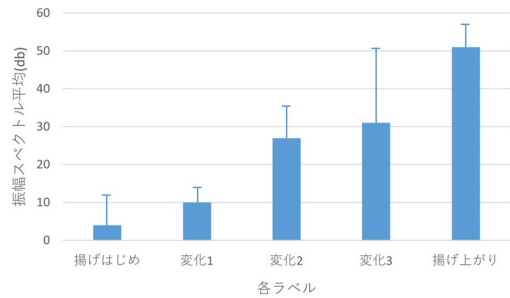


図 4: 各ラベルごとの平均振幅スペクトルの推移

## 4. 提案手法

本節では音響信号を用いた揚げ具合推定の手法について述べる。

### 4.1 データ前処理

本実験で音声データは安静な屋内キッチンから取得したが、マイクの接触、会話、音声拡張子変換時圧縮による雑音を含めている。図 5 はバンドパスフィルタ [8] を適応前後のフーリエ変換の結果を示す。適応前のフーリエ変換から、3000Hz 以下の低い周波数を除いてスペクトラムのピークが 5000Hz 前後にある。そのため今回は音声波形の 2000Hz 8000Hz のバンドパスフィルタを適応し調節した。

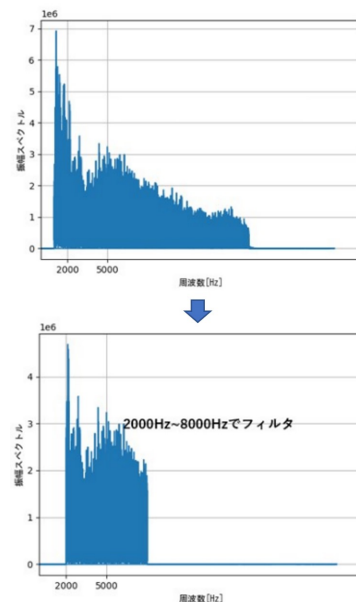


図 5: バンドパスフィルタ適応結果

### 4.2 揚げ具合の特徴抽出

音響特徴量にはメル周波数ケプストラム係数 (MFCC)[9] を使用した。MFCC は音声認識や楽器の識別などによく使用される特徴量であり、メル尺度・メル周波数・メルスペクトルを使用している。メル尺度とは人間の聴覚に基づ

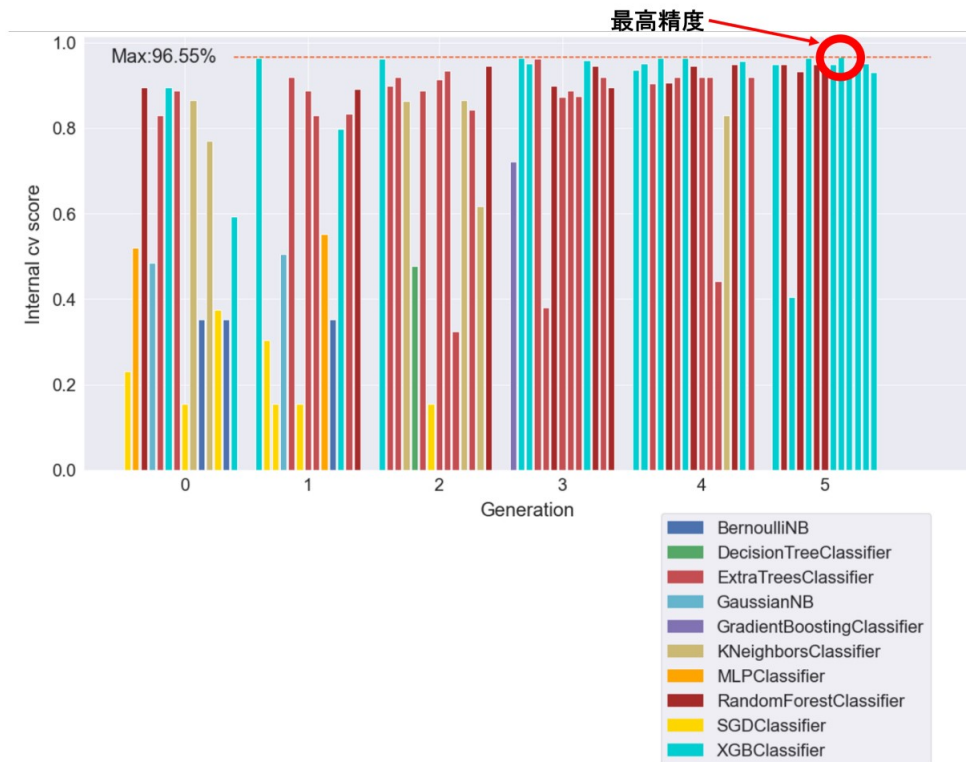


図 6: 分類器ごとの精度の移り変わり

いた尺度であり、人間の聴覚は周波数の低い音に対して敏感に、周波数の高い音に関しては鈍感であるという性質がある。メル周波数とは人間の聴覚に即した周波数スケールのことであり、メル尺度を使用して作ることができる。そこからさらにメル周波数のスケールを作成し、メルフィルタバンクにかけることによってメル周波数スペクトル係数を表すことができる。

メルフィルタバンクとはスペクトルにかける三角窓のフィルタである。また、今回この音響特徴量を使用した理由としては MFCC が人間の聴覚に基づいた尺度で得られる特徴量だからである。唐揚げを揚げる際に出る音は人間の耳で感じ取れる音の変化であるためにこの特徴量を採用した。特徴量として扱う MFCC は 1 次から 20 次の 20 個係数として、メルフィルタバンクも 20 チャンネルと設定した。

#### 4.3 機械学習手法

機械学習では教師あり学習を使用した。また、使用した音響特徴量は前節で述べた MFCC であり、機械学習には Python のツールである TPOT を使用した [10]。TPOT とは遺伝的プログラミングを使用して機械学習パイプラインを最適化する自動機械学習ツール (AutoML) である。TPOT が実施するのは変換器と予測器の生成である。そして、変換器と予測器を直列に繋げるパイプラインを作る。

TPOT では特徴量工学・メタ学習・アーキテクチャサ

チなどの構成要素を組み合わせることによってモデル選択とハイパーパラメータ最適化の難しさを克服することが可能になっている。また、特徴処理・モデル・ハイパーパラメータを考慮し、与えられた分類や回帰のタスクに対して最適なパイプラインを探索する。処理モジュールは特徴選択モジュールおよび生成モジュールと連動して動作し、特徴構築ブロックはカーネルトリックや次元削減処理を行う。そして、最終的にパイプラインは遺伝的プログラミングによって最適化される。パラメータとして、変異するジェネレーションを 5 とし、ポピュレーションサイズを 15 と設定し、5 分割交差検証を用いて唐揚げ揚げ具合の各段階の予測推定を行った。

## 5. 評価結果

実験データに対する機械学習分類の結果と考察を述べる。

### 5.1 TPOT による機械学習モデル

機械学習による分析の結果、最も精度が高く分類することができたのは、XGBClassifier (勾配ブースティング) であり、5 分割交差検証の精度は 96.55% まで達した。また、使用された分類器は 10 種類であり、パラメータ調節を含んだ総合の試行回数は 84 回である。分類器ごとの精度の変化は図 6 のようになる。横軸である Generation が増えると同時に交差と変異が働き、パラメータを自動的に高い精度の方に調節する。Generation が 5 の時に XGBClassifier

表 2: 各クラスにおける評価結果

評価尺度	その他	揚げはじめ	変化 1	変化 2	変化 3	揚げ上がり	平均	加重平均
Recall	98.53%	95.71%	98.66%	95.96%	98.58%	83.33%	95.13%	97.15%
Specificity	99.79%	99.63%	98.14%	99.52%	99.40%	99.80%	99.38%	99.15%
Precision	97.10%	98.67%	95.45%	98.17%	97.66%	88.24%	95.88%	97.17%
Accuracy	98.53%	95.71%	98.66%	95.96%	98.58%	83.33%	95.13%	97.15%
F1-score	97.81%	97.17%	97.03%	97.05%	98.12%	85.71%	95.48%	97.14%

の最高精度が検出された様子を表している。

## 5.2 モデル分類結果

構成された Xgboost モデルの結果からの混同行列は図 7 のようになり、各クラスにおける評価結果は表 2 のようになる。表 2 からラベルごとの評価に着目すると、揚げ上がり以外の精度が約 95%以上とかなり、高い精度で分類することができているとわかる。また揚げ上がりの中でも特異度がかなり高く、正しいデータを間違えて誤分類することが少ないことがわかる。

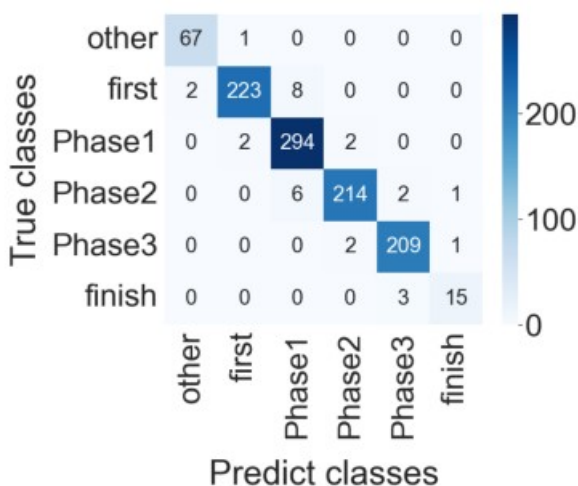


図 7: モデルの混同行列

## 5.3 考察

当初の機械学習を使用せずに、振幅スペクトルの平均で判別を行う手法について検討していたが、表 1 のように解析に使用したデータは高温・適温・低温・量多め・量少なめと様々な種類の揚げ方を使用しており、それぞれの特徴が強く表れていた。例えば、高温・量少なめの場合揚げ上がる速度が速く、低温・量多めの場合揚げ終わる速度が遅い。また、量多めだと発生する音の振幅スペクトルが小さいという特徴があった。その中でも特に、量多めで発生する振幅スペクトルが小さいという特徴は平均振幅スペクトルを閾値に設定することをより困難にしていた。そこで振幅スペクトルのみでは解析をすることができないとわかり、機械学習を使用することにした。機械学習を使用す

ることによって予測精度が 96.55%、F1-score が 97.14% と高い精度で分類できることがわかった。しかし、問題点としては最終ラベルの揚げ上がりのデータ数が少なく、他のラベルの精度と比べてしまうと精度が低くなってしまっている。その解決策としては、今回のようにラベルを経過的に分類させることによって揚げ上がりのタイミングの検知が可能になると考えた。混同行列から見て誤分類は変化 1 の周りとなげ上がりに分布している。ラベル変化 1 の判断が難しいと考えられる。唐揚げの色変化、泡の量に変化がなく、ラベル変化 1 の定義の明確化が求められる。また揚げ上がりの判定について、データ数が少ないにもかかわらず割と高い精度で分類が可能だと分かる。音声による揚げ具合の推定も実用性はあると考えられる。

## 6. 結論

### 6.1 まとめ

本研究では揚げ物を揚げ具合を揚げはじめ・変化 1・変化 2・変化 3・揚げ上がり 5 つのラベルと定義し、実際の機械学習によるモデルの有効性を検証した。揚げ物に関して、料理初心者が最も難しいと感じている点として中まで火が通っているのかどうか分からないということが挙げられる。しかし本論文のように機械学習を使用することによって F1-score 97.14% の高い精度で唐揚げの揚げ上りを判断できることがわかった。従来のレシピなどでは目安の油の温度や時間が掲載されているが、本論文の手法であれば自由な温度で揚げ上りを自動的に取得することが可能である。このような調節のいらぬ手法は料理初心者にとって必要な機能であり、初心者支援として機能すると考えられる。

### 6.2 今後の展望

本論文では機械学習を用いた手法を提案した。今回のような機械学習による判別を実際を使用することを想定すると、窓 10 秒がラベルの境界線に存在する場合の誤判別の可能性がある。よって、判別結果を 1 秒に 1 回返すことを想定した場合、揚げ上がりの判別が 5 回生じたタイミングを揚げ上がりとすることによって精度が安定すると考える。この技術をスマートフォンや調理器具に組み込むことができれば、誰でも楽に唐揚げを作ることができる。今後の展望としてこの機械学習による判別をデバイスや鍋など

に埋め込む方法を検討していきたい。そのためには揚げあがったときに通知する手段や調理環境に影響されないようにする必要がある。特に油を使用する環境では油が跳ねることや高温な環境に耐えられるものを使用する必要があると考えられる。

## 参考文献

- [1] 「食育に関する意識調査の結果-農林水産省」  
available from (<https://www.maff.go.jp/j/press/syouan/hyoji/210331.html>)
- [2] Kido Yuta, Mizumoto Teruhiro, Suwa Hirohiko, Arakawa Yutaka and Yasumoto Keiichi. "A Cooking Support System for Seasoning with Smart Cruet." International Conference on Human-Computer Interaction. Springer, Cham, 2019.
- [3] 加藤岳大, 横窪安奈, ロペズギヨーム. SynCook: 動画メタデータと加速度センサを用いたレシピ動画進行度自動同期システム. WISS, 2020.
- [4] Voong KY, Norton AB, Mills TB and Norton IT, Characterisation of deep-fried batter and breaded coatings. Food structure, 16: 43-49, 2018.
- [5] Pedreschi F., Hernández P., Figueroa C., and Moyano P. Modeling water loss during frying of potato slices. International journal of food properties, 8(2), 289-299, 2005.
- [6] Yusaku Korematsu, Daisuke Saito, and Nobuaki Mine-matsu. Cooking state recognition based on acoustic event detection. In Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities, CEA ' 19, p. 41-44. Association for Computing Machinery, 2019.
- [7] 温度計 揚げもの用温度計クックサーモ 5495B (クロム)  
available from ([https://shop.tanita.co.jp/shop/g/\\_TDT495ST/](https://shop.tanita.co.jp/shop/g/_TDT495ST/))
- [8] Yost William A, Pastore M Torben, Dorman and Michael F. Sound source localization is a multisystem process. Acoustical science and technology, 41(1), 113-120, 2020.
- [9] Zheng Fang, Guoliang Zhang and Zhanjiang Song. Comparison of different implementations of MFCC. Journal of Computer science and Technology, 16(6), 582-589, 2001.
- [10] TPOT a Python Automated Machine Learning tool  
available from (<http://epistasislab.github.io/tpot/>)