1. **First experiment:**

**Title**: Two-stage heart arrhythmia ensemble classifier

**Hypothesis**: The classification of heart arrhythmia performed in two-stage using the proposed ensemble classifier compared to one-stage would improve the classification of the Arrhythmias present in the data.

**Experimental set up**: The features present in the dataset missing values were imputed using data imputation techniques, and Isolation Forest was used on the healthy population as a method to remove outliers and under sample the healthy population. The XGBoost classifier was fitted with 60% data using cross validation for the first stage that only classifies normal/abnormal. Then the abnormal cases were fed into the second classifier with 60/40 split and the results were recorded. No oversampling technique was used for this experiment.

**ML Technique**: Isolation Forest for under sampling, XGBoost for the classifier

**Metrics** : F1-score, Precision, Recall

**Statistical Method**: Cross-validation

**Results**: one-stage vs two-stage

One-stage Results

|  |  |  |  |
| --- | --- | --- | --- |
| Label | Recall | Precision | F1Score |
| Normal | 0.990 | 0.960 | 0.975 |
| AF | 0.390 | 0.720 | 0.506 |
| Arr | 0.020 | 0.120 | 0.034 |

Two-stage Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Label | Recall AF | Recall Arr | Precision | F1Score 0 | F1Score 1 | Total F1Score |
| First-stage | 0.90 |  | 0.71 | 0.98 (Healthy) | 0.56 (Unhealthy) | 0.77 |
| Second-stage | 0.54 |  | 0.53 | 0.14 (Arr) | 0.93 (AF) | 0.53 |

**Conclusion**: Classification of Arr and AF is more efficient when performed in a two-stage manner.

**Further work**: The two-stage classifier that classified AF-Arr, although showed improvement in the classification of Arrhythmias, but improvements can still be made to improve the classification of AF- Arr only, since the algorithm still struggles to distinguish between the two diseases. This can be seen from the low F1 score for Arr classification.

1. **Second Experiment**

**Title**: Improvement of the AF/Arr classifier 🡪 2nd stage only

**Hypothesis**: The combination of outlier removal, oversampling method as well as a voting classifier created from XGBoost, Naïve Bayes classifier and Stochastic Gradient Descent would have a significant improvement on the classification of AF\Arr for the classifier.

**Expected Contribution**: the classifier is improved significantly when outlier removal and oversampling is used.

**Experimental set up**: comparison of the following options:

**Two algorithms**  🡪 XGBoost (XG) and Voting Classifier (VC)

**Outlier removal methods** 🡪 Isolation Forest (IF) and Local outlier Factor (LOF), No (N)

**Oversampling** 🡪 over sampler (OS) and SMOTE, No (N)

Uk BioBank dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Outlier Removal** | **Over-Sampling** | **Precision** | **Recall** | **F1 for AF** | **F1 for Arr** | **F1 total** |
| XG | No | No | 0.79 | 0.52 | 0.74 | 0.46 | 0.60 |
| XG | IF | No | 0.77 | 0.90 | 0.69 | 0.18 | 0.54 |
| XG | LOF | No | 0.25 | 0.50 | 0.67 | 0 | 0.33 |
| XG | No | OS | 0.22 | 0.40 | 0.57 | 0 | 0.29 |
| XG | No | SMOTE | 0.77 | 0.55 | 0.69 | 0.18 | 0.44 |
| XG | IF | OS | 0.79 | 0.52 | 0.74 | 0.46 | 0.60 |
| XG | IF | SMOTE | 0.60 | 0.52 | 0.74 | 0.46 | 0.60 |
| XG | LOF | OS | 0.50 | 0.50 | 0.64 | 0.17 | 0.40 |
| XG | LOF | SMOTE | 0.60 | 0.55 | 0.67 | 0.31 | 0.49 |
| VC | No | No | 0.78 | 0.60 | 0.71 | 0.33 | 0.52 |
| VC | IF | No | 0.67 | 0.65 | 0.59 | 0.70 | 0.64 |
| VC | LOF | No | 0.78 | 0.60 | 0.71 | 0.33 | 0.52 |
| VC | No | OS | 0.50 | 0.50 | 0.44 | 0.55 | 0.49 |
| VC | No | SMOTE | 0.78 | 0.60 | 0.71 | 0.33 | 0.52 |
| VC | IF | OS | 0.70 | 0.65 | 0.72 | 0.53 | 0.63 |
| VC | IF | SMOTE | 0.86 | 0.85 | 0.87 | 0.86 | 0.87 |
| VC | LOF | OS | 0.56 | 0.55 | 0.53 | 0.57 | 0.55 |
| VC | LOF | SMOTE | 0.40 | 0.40 | 0.33 | 0.45 | 0.39 |

UCI dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Outlier Removal** | **Over-Sampling** | **Precision** | **Recall** | **F1 total** | **Accuracy** |
| XG | No | No | 0.59 | 0.64 | 0.61 | 0.76 |
| XG | IF | No | 0.45 | 0.42 | 0.41 | 0.62 |
| XG | LOF | No | 0.60 | 0.51 | 0.52 | 0.62 |
| XG | No | OS | 0.30 | 0.29 | 0.29 | 0.48 |
| XG | No | SMOTE | 0.50 | 0.47 | 0.47 | 0.62 |
| XG | IF | OS | 0.22 | 0.27 | 0.24 | 0.52 |
| XG | IF | SMOTE | 0.40 | 0.36 | 0.36 | 0.52 |
| XG | LOF | OS | 0.41 | 0.33 | 0.35 | 0.52 |
| XG | LOF | SMOTE | 0.35 | 0.35 | 0.35 | 0.52 |
| VC | No | No | 0.67 | 0.69 | 0.67 | 0.71 |
| VC | IF | No | 0.56 | 0.57 | 0.56 | 0.62 |
| VC | LOF | No | 0.57 | 0.56 | 0.54 | 0.67 |
| VC | No | OS | 0.72 | 0.64 | 0.65 | 0.76 |
| VC | No | SMOTE | 0.72 | 0.69 | 0.69 | 0.81 |
| VC | IF | OS | 0.52 | 0.49 | 0.49 | 0.71 |
| VC | IF | SMOTE | 0.60 | 0.57 | 0.57 | 0.76 |
| VC | LOF | OS | 0.52 | 0.58 | 0.55 | 0.76 |
| VC | LOF | SMOTE | 0.56 | 0.58 | 0.56 | 0.76 |

**ML Technique**: The final technique selected include Isolation Forest for outlier removal, SMOTE for oversampling, soft Voting classifier for the classifier (XGBoost, Naïve Bayes, Stochastic Gradient Descent) based on the results.

**Metrics** : F1-score, specificity (precision), sensitivity (recall)

**Statistical Method**: Cross validation

**Results**: the best combination is the voting classifier plus isolation forest plus smote

**Conclusion**: Classification of Arr and AF is more efficient when performed using the combination of VC + IF + SM.

**Further work**: Going back to the two-stage classifier, we can now utilise this in our second stage and come up with the best combination for our two-stage classifier.

1. **Third Experiment:**

**Title:** Two stage classifier improvement method combinations.

**Hypothesis**: The addition of the improvement for the second stage classifier as shown in previous experiment should have a substantial effect on the two-stage classifier overall performance. The second stage classifier that will be used for the experiment is the combination of **VC + IF + SM.**

**Experimental Set up**: Combination of the following for the first stage.

**Algorithms** 🡪 XG (XGBoost) and VC (Voting Classifier)

**Outlier Removal** 🡪 IF (Isolation Forest) and LOF (Local outlier Factor) , No (N)

**Oversampling** 🡪 OS (Oversampler) and SMOTE, No(N)

**ML Technique:** XGBoost and Voting Classifier, Isolation Forest, and Local Outlier Factor, Oversampler and SMOTE for the oversampling technique

**Metrics** : F1-score, specificity (precision), sensitivity (recall)

**Statistical Method**: Cross validation

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stage one | P | R | Healthy F1 | Unhealthy F1 | T F1 | Stage Two | P | R | AF F1 | Arr F1 | T F1 |
| XG | 0.58 | 0.52 | 0.84 | 0.09 | 0.47 | **VC + IF + SM** | 0.52 | 0.52 | 0.92 | 0.12 | 0.52 |
| XG + IF | 0.66 | 0.59 | 0.84 | 0.36 | 0.60 | **VC + IF + SM** | 0.52 | 0.55 | 0.61 | 0.18 | 0.40 |
| XG + LOF | 0.71 | 0.52 | 0.85 | 0.08 | 0.47 | **VC + IF + SM** | 0.55 | 0.55 | 0.18 | 0.18 | 0.18 |
| XG + SM | 0.55 | 0.51 | 0.84 | 0.06 | 0.45 | **VC + IF + SM** | 0.51 | 0.52 | 0.70 | 0.16 | 0.43 |
| XG + OS | 0.62 | 0.53 | 0.84 | 0.17 | 0.51 | **VC + IF + SM** | 0.54 | 0.60 | 0.75 | 0.22 | 0.48 |
| XG + IF + SM | 0.55 | 0.53 | 0.82 | 0.21 | 0.51 | **VC + IF + SM** | 0.55 | 0.64 | 0.76 | 0.24 | 0.50 |
| XG + IF + OS | 0.56 | 0.54 | 0.81 | 0.26 | 0.54 | **VC + IF + SM** | 0.55 | 0.65 | 0.57 | 0.23 | 0.40 |
| XG + LOF + SM | 0.65 | 0.52 | 0.85 | 0.10 | 0.47 | **VC + IF + SM** | 0.54 | 0.62 | 0.69 | 0.22 | 0.46 |
| XG + LOF + OS | 0.59 | 0.52 | 0.84 | 0.13 | 0.48 | **VC + IF + SM** | 0.52 | 0.54 | 0.41 | 0.18 | 0.29 |
| VC | 0.54 | 0.50 | 0.85 | 0.01 | 0.43 | **VC + IF + SM** | 0.42 | 0.44 | 0.75 | 0.10 | 0.42 |
| VC + IF | 0.57 | 0.56 | 0.80 | 0.32 | 0.56 | **VC + IF + SM** | 0.55 | 0.53 | 0.93 | 0.14 | 0.54 |
| VC + LOF | 0.64 | 0.51 | 0.85 | 0.08 | 0.46 | **VC + IF + SM** | 0.60 | 0.54 | 0.94 | 0.17 | 0.55 |
| VC + SM | 0.62 | 0.65 | 0.74 | 0.50 | 0.62 | **VC + IF + SM** | 0.49 | 0.47 | 0.61 | 0.14 | 0.38 |
| VC + OS | 0.64 | 0.64 | 0.81 | 0.46 | 0.64 | **VC + IF + SM** | 0.58 | 0.67 | 0.85 | 0.30 | 0.58 |
| VC + IF + SM | 0.57 | 0.55 | 0.81 | 0.30 | 0.55 | **VC + IF + SM** | 0.53 | 0.55 | 0.86 | 0.18 | 0.52 |
| VC + IF + OS | 0.54 | 0.53 | 0.78 | 0.28 | 0.53 | **VC + IF + SM** | 0.50 | 0.50 | 0.50 | 0.16 | 0.33 |
| VC + LOF + SM | 0.59 | 0.57 | 0.81 | 0.34 | 0.57 | **VC + IF + SM** | 0.52 | 0.56 | 0.72 | 0.19 | 0.46 |
| VC + LOF + OS | 0.65 | 0.57 | 0.84 | 0.29 | 0.57 | **VC + IF + SM** | 0.51 | 0.52 | 0.86 | 0.15 | 0.50 |

UK BioBank dataset

UCI dataset

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stage one | P | R | Healthy F1 | Unhealthy F1 | T F1 | Acc | Stage Two | P | R | T F1 | Acc |
| XG | 0.72 | 0.72 | 0.74 | 0.71 | 0.72 | 0.73 | **VC** | 0.60 | 0.56 | 0.57 | 0.76 |
| XG + IF | 0.72 | 0.71 | 0.68 | 0.72 | 0.70 | 0.70 | **VC** | 0.62 | 0.59 | 0.58 | 0.76 |
| XG + LOF | 0.74 | 0.74 | 0.74 | 0.73 | 0.74 | 0.74 | **VC** | 0.60 | 0.54 | 0.55 | 0.76 |
| XG + SM | 0.73 | 0.73 | 0.73 | 0.72 | 0.73 | 0.73 | **VC** | 0.71 | 0.73 | 0.72 | 0.76 |
| XG + OS | 0.69 | 0.69 | 0.72 | 0.66 | 0.69 | 0.69 | **VC** | 0.65 | 0.53 | 0.55 | 0.71 |
| XG + IF + SM | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | **VC** | 0.64 | 0.58 | 0.57 | 0.71 |
| XG + IF + OS | 0.76 | 0.76 | 0.77 | 0.74 | 0.76 | 0.76 | **VC** | 0.65 | 0.61 | 0.60 | 0.74 |
| XG + LOF + SM | 0.68 | 0.68 | 0.69 | 0.67 | 0.68 | 0.68 | **VC** | 0.58 | 0.56 | 0.55 | 0.71 |
| XG + LOF + OS | 0.76 | 0.76 | 0.77 | 0.74 | 0.76 | 0.76 | **VC** | 0.62 | 0.56 | 0.57 | 0.76 |
| VC | 0.70 | 0.70 | 0.70 | 0.73 | 0.67 | 0.70 | **VC** | 0.64 | 0.56 | 0.56 | 0.76 |
| VC + IF | 0.70 | 0.70 | 0.72 | 0.68 | 0.70 | 0.70 | **VC** | 0.63 | 0.59 | 0.58 | 0.76 |
| VC + LOF | 0.74 | 0.74 | 0.74 | 0.73 | 0.74 | 0.74 | **VC** | 0.65 | 0.56 | 0.57 | 0.76 |
| VC + SM | 0.71 | 0.71 | 0.71 | 0.69 | 0.71 | 0.71 | **VC** | 0.63 | 0.53 | 0.55 | 0.74 |
| VC + OS | 0.71 | 0.71 | 0.74 | 0.68 | 0.75 | 0.75 | **VC** | 0.62 | 0.56 | 0.57 | 0.76 |
| VC + IF + SM | 0.75 | 0.75 | 0.75 | 0.74 | 0.75 | 0.75 | **VC** | 0.60 | 0.61 | 0.59 | 0.74 |
| VC + IF + OS | 0.71 | 0.71 | 0.74 | 0.68 | 0.71 | 0.71 | **VC** | 0.62 | 0.56 | 0.57 | 0.76 |
| VC + LOF + SM | 0.76 | 0.74 | 0.79 | 0.68 | 0.74 | 0.75 | **VC** | 0.60 | 0.51 | 0.54 | 0.71 |
| VC + LOF + OS | 0.77 | 0.76 | 0.80 | 0.73 | 0.77 |  | **VC** | 0.68 | 0.64 | 0.63 | 0.79 |

**Conclusion**: surprisingly, the combination of VC + OS for the first stage and VC + IF + SM for the second stage gave the best results for the two-stage classifier among other combinations. This experiment is still in progress and patterns are being discovered, the results above are subject to change. The results of this experiment are being compared to the one stage classifier which is shown in the next experiment.

1. **fourth experiment:**

**Title**: One stage classifier combinations:

**Hypothesis**: comparison of the two-stage classifier performance with the one-stage. The hypothesis is that the two-stage classifier should have a higher f1-score for the classification of diseases and hence perform better.

**Experimental Set up**: The combination is set out below, all performed in one-stage.

**Oversampling** 🡪 OS and SMOTE, No(N)

**ML Technique:** XGBoost and Voting Classifier, Isolation Forest, and Local Outlier Factor, Oversampler and SMOTE for the oversampling technique

**Metrics** : F1-score, specificity (precision), sensitivity (recall)

**Statistical Method**: Cross Validation

UK BioBank results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Method | Precision | Recall | Healthy F1 | AF F1 | Arr F1 | Total F1 |
| XG | 0.44 | 0.34 | 0.85 | 0.05 | 0.0 | 0.30 |
| XG + IF | 0.39 | 0.37 | 0.83 | 0.25 | 0.0 | 0.36 |
| XG + LOF | 0.47 | 0.34 | 0.85 | 0.02 | 0.0 | 0.29 |
| XG + SM | 0.47 | 0.34 | 0.85 | 0.02 | 0.0 | 0.29 |
| XG + OS | 0.45 | 0.36 | 0.85 | 0.16 | 0.0 | 0.34 |
| XG + IF + SM | 0.41 | 0.38 | 0.84 | 0.28 | 0.0 | 0.37 |
| XG + IF + OS | 0.37 | 0.37 | 0.81 | 0.29 | 0.0 | 0.37 |
| XG + LOF + SM | 0.58 | 0.34 | 0.85 | 0.02 | 0.0 | 0.29 |
| XG + LOF + OS | 0.38 | 0.35 | 0.84 | 0.12 | 0.0 | 0.32 |
| VC | 0.58 | 0.34 | 0.85 | 0.01 | 0.0 | 0.29 |
| VC + IF | 0.39 | 0.38 | 0.83 | 0.29 | 0.0 | 0.37 |
| VC + LOF | 0.25 | 0.33 | 0.85 | 0.0 | 0.0 | 0.28 |
| VC + SM | 0.44 | 0.37 | 0.82 | 0.17 | 0.04 | 0.35 |
| VC + OS | 0.43 | 0.43 | 0.83 | 0.47 | 0.0 | 0.43 |
| VC + IF + SM | 0.38 | 0.37 | 0.81 | 0.23 | 0.05 | 0.36 |
| VC + IF + OS | 0.45 | 0.38 | 0.81 | 0.26 | 0.10 | 0.39 |

UCI Results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Precision | Recall | Total F1 | Accuracy |
| XG | 0.52 | 0.48 | 0.48 | 0.71 |
| XG + IF | 0.40 | 0.42 | 0.39 | 0.59 |
| XG + LOF | 0.44 | 0.34 | 0.37 | 0.66 |
| XG + SM | 0.44 | 0.44 | 0.42 | 0.70 |
| XG + OS | 0.37 | 0.33 | 0.33 | 0.62 |
| XG + IF + SM | 0.34 | 0.38 | 0.34 | 0.62 |
| XG + IF + OS | 0.41 | 0.36 | 0.36 | 0.60 |
| XG + LOF + SM | 0.35 | 0.39 | 0.36 | 0.59 |
| XG + LOF + OS | 0.47 | 0.35 | 0.38 | 0.60 |
| VC | 0.51 | 0.51 | 0.50 | 0.69 |
| VC + IF | 0.53 | 0.45 | 0.45 | 0.66 |
| VC + LOF | 0.54 | 0.41 | 0.44 | 0.68 |
| VC + SM | 0.58 | 0.49 | 0.51 | 0.73 |
| VC + OS | 0.46 | 0.42 | 0.43 | 0.68 |
| VC + IF + SM | 0.45 | 0.42 | 0.42 | 0.69 |
| VC + IF + OS | 0.41 | 0.36 | 0.36 | 0.60 |

**Conclusion**: surprisingly for the one stage classifier the VC + OS combination also performed the best. But as it can be seen the classifier is unable to classify any Arrythmia cases when this is done in a one stage manner.

* Next points to consider would be whether to repeat these experiments with the UCI publicly available datasets to have a better comparison

Trying the above experiments with UCI dataset: