Muhammad Mahad Sheikh

21i-1239

Assignment#2

Feature Selection Process:

Selecting and combining data: The data selection process involves choosing columns with modality 01 and emotions 03 or 04. Once these columns are identified, the data representing facial recognition is selected, and all relevant data is stored and combined in dataframes.

Initialize Population: The population for the genetic algorithm is initialized by randomly assigning 0 and 1 to represent the absence or presence of a feature.

Fitness Function: The fitness function evaluates the performance of each chromosome (binary string representing selected features) using a neural network model. This model, built using the Keras Sequential API, consists of several dense layers with ReLU activation functions, followed by a softmax activation function in the output layer. The model is compiled with appropriate loss function, optimizer, and evaluation metric. EarlyStopping callback is used to prevent overfitting. After training, the model's accuracy on the testing dataset is computed and returned as the fitness score.

Crossover Function: The crossover function mimics the process of genetic recombination by randomly selecting a crossover point and swapping segments of two parent chromosomes to create two offspring chromosomes. This process introduces diversity in the population.

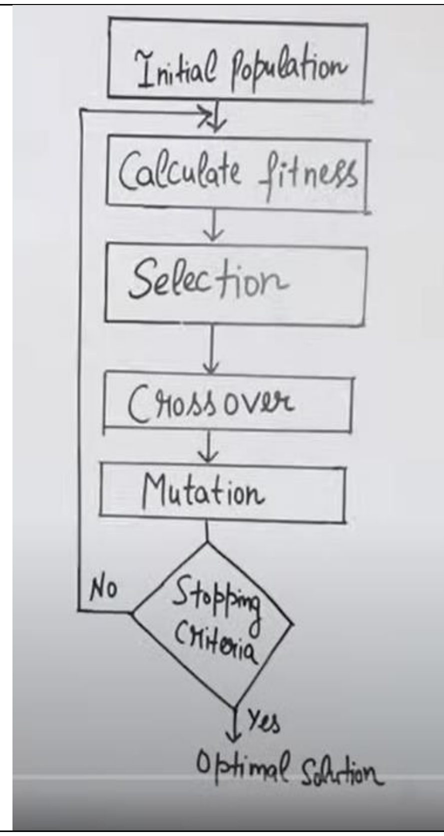
Mutation Function: The mutation function introduces random changes in the chromosomes by flipping individual genes with a certain probability. This randomness helps explore new solutions in the search space.

Genetic Algorithm: The genetic algorithm orchestrates the feature selection process. It iteratively evaluates the fitness of each chromosome in the population, selects parent chromosomes based on fitness scores, performs crossover and mutation operations to create new offspring chromosomes, and updates the population by replacing less fit chromosomes with mutated offspring. The best solution found along with its fitness score is tracked throughout the process. This iterative process aims to converge towards an optimal solution for feature selection.

Overall, this feature selection approach combines elements of evolutionary algorithms with neural network models to efficiently identify a subset of features that optimize classification accuracy.

1. **Explain the chromosome representation and genetic operators used in the GA.**

Chromosome: This is a binary string, where each bit is representing a feature whether its selected or not. Off course 1 being the selected and 0 being not. e.g) 1001 means first and fourth feature is selected and 2nd and 3rd feature is not.

Initialization: Population initialized randomly with 0s and 1s representing absence/presence of features. **Fitness Function**: Evaluates chromosome performance using a neural network, measuring accuracy of model trained on selected features. **Selection:** Parent chromosomes chosen probabilistically based on fitness scores; higher fitness = higher chance of selection. **Crossover:** Recombination between selected parents generates offspring, combining features from both. **Mutation:** Introduces random changes in chromosomes, flipping genes with a certain probability to explore new solutions. **Stopping Criteria:** The stopping criteria, particularly for the EarlyStopping callback, is not explicitly mentioned in the provided information. Typically, EarlyStopping monitors a specified metric (e.g., validation loss) during training and stops training if the metric does not improve after a certain number of epochs (patience). However, the exact stopping criteria, such as the maximum number of epochs or a predefined threshold for improvement, are not specified in the given report.	

2. Present the performance evaluation results for the final model.

Population Size: 10 Generation: 100 ***Note: Have selected the best fitness solution for each generation.***	Accuracy with all features: 0.6320563554763794 Accuracy with selected features: 0.7048873901367188 Feature selection improved classification accuracy.

1. Generation 18 Accuracies: [0.8254721760749817]
2. Generation 84 Accuracies: [0.7690214514732361]
3. Generation 56 Accuracies: [0.769981861114502]
4. Generation 17 Accuracies: [0.7813466787338257]
5. Generation 76 Accuracies: [0.7823604941368103]
6. Generation 57 Accuracies: [0.8015686869621277]
7. Generation 24 Accuracies: [0.8079180717468262]
8. Generation 38 Accuracies: [0.8079714179039001]
9. Generation 46 Accuracies: [0.807996392250061]
10. Generation 2 Accuracies: [0.6568669080734253]
11. Generation 28 Accuracies: [0.7860420346260071]
12. Generation 20 Accuracies: [0.786789059638977]
13. Generation 85 Accuracies: [0.7923380732536316]
14. Generation 53 Accuracies: [0.7739301919937134]
15. Generation 82 Accuracies: [0.7739301919937134]
16. Generation 74 Accuracies: [0.7594707012176514]
17. Generation 50 Accuracies: [0.7594707012176514]
18. Generation 34 Accuracies: [0.6745278239250183]
19. Generation 69 Accuracies: [0.7317255139350891]
20. Generation 68 Accuracies: [0.7333261966705322]
21. Generation 39 Accuracies: [0.6858392953872681]
22. Generation 22 Accuracies: [0.6341372132301331]
23. Generation 91 Accuracies: [0.605218231678009]
24. Generation 94 Accuracies: [0.7442642450332642]
25. Generation 95 Accuracies: [0.6794365644454956]
26. Generation 96 Accuracies: [0.7129442095756531]
27. Generation 93 Accuracies: [0.5720840692520142]
28. Generation 1 Accuracies: [0.7387151718139648]
29. Generation 3 Accuracies: [0.7242556810379028]
30. Generation 6 Accuracies: [0.7247892618179321]
31. Generation 29 Accuracies: [0.7230284810066223]
32. Generation 62 Accuracies: [0.7223882079124451]
33. Generation 97 Accuracies: [0.6995518207550049]
34. Generation 99 Accuracies: [0.6947497725486755]
35. Generation 98 Accuracies: [0.6959235668182373]
36. Generation 63 Accuracies: [0.697844386100769]
37. Generation 23 Accuracies: [0.7001920938491821]
38. Generation 65 Accuracies: [0.7030733227729797]
39. Generation 92 Accuracies: [0.7034468054771423]
40. Generation 41 Accuracies: [0.7050474882125854]
41. Generation 10 Accuracies: [0.7052075266838074]
42. Generation 15 Accuracies: [0.7046740055084229]
43. Generation 11 Accuracies: [0.6783694624900818]
44. Generation 40 Accuracies: [0.6386724710464478]
45. Generation 42 Accuracies: [0.6039376854896545]
46. Generation 91 Accuracies: [0.605218231678009]
47. Generation 44 Accuracies: [0.6149290204048157]
48. Generation 77 Accuracies: [0.6151424646377563]
49. Generation 45 Accuracies: [0.6582542061805725]
50. Generation 30 Accuracies: [0.7398356795310974]
51. Generation 31 Accuracies: [0.5913456678390503]
52. Generation 33 Accuracies: [0.7489061951637268]
53. Generation 54 Accuracies: [0.7460783123970032]
54. Generation 43 Accuracies: [0.7458649277687073]
55. Generation 71 Accuracies: [0.745384693145752]
56. Generation 61 Accuracies: [0.7555223703384399]
57. Generation 25 Accuracies: [0.7582435011863708]
58. Generation 81 Accuracies: [0.757176399230957]
59. Generation 9 Accuracies: [0.6874933242797852]
60. Generation 88 Accuracies: [0.7195603251457214]
61. Generation 89 Accuracies: [0.7149183750152588]
62. Generation 72 Accuracies: [0.6861060857772827]
63. Generation 90 Accuracies: [0.6862127780914307]
64. Generation 21 Accuracies: [0.6745811700820923]
65. Generation 7 Accuracies: [0.7188133597373962]
66. Generation 37 Accuracies: [0.5581581592559814]
67. Generation 55 Accuracies: [0.6897876262664795]
68. Generation 14 Accuracies: [0.6879201531410217]
69. Generation 35 Accuracies: [0.728257417678833]
70. Generation 87 Accuracies: [0.6734073162078857]
71. Generation 59 Accuracies: [0.7278838753700256]
72. Generation 78 Accuracies: [0.7290043830871582]
73. Generation 64 Accuracies: [0.6182371377944946]
74. Generation 66 Accuracies: [0.7410094738006592]
75. Generation 67 Accuracies: [0.588197648525238]
76. Generation 19 Accuracies: [0.690961480140686]
77. Generation 13 Accuracies: [0.71059650182724]
78. Generation 8 Accuracies: [0.7382349967956543]
79. Generation 79 Accuracies: [0.7684345245361328]
80. Generation 80 Accuracies: [0.7327393293380737]
81. Generation 75 Accuracies: [0.7076619267463684]
82. Generation 73 Accuracies: [0.7203606963157654]
83. Generation 70 Accuracies: [0.5644541382789612]
84. Generation 47 Accuracies: [0.583182156085968]
85. Generation 12 Accuracies: [0.7441041469573975]
86. Generation 86 Accuracies: [0.7411162257194519]
87. Generation 48 Accuracies: [0.7177462577819824]
88. Generation 5 Accuracies: [0.7510938048362732]
89. Generation 36 Accuracies: [0.7485327124595642]
90. Generation 60 Accuracies: [0.7440508008003235]
91. Generation 32 Accuracies: [0.7329527139663696]
92. Generation 51 Accuracies: [0.6361647844314575]
93. Generation 58 Accuracies: [0.7619784474372864]
94. Generation 26 Accuracies: [0.5471667647361755]
95. Generation 49 Accuracies: [0.4916764497756958]
96. Generation 16 Accuracies: [0.7171059846878052]
97. Generation 52 Accuracies: [0.766620397567749]
98. Generation 4 Accuracies: [0.7056877613067627]
99. Generation 27 Accuracies: [0.6623092293739319]
100. Generation 100 Accuracies: [0.6162629127502441]

Population Size: 10 Generation: 10 ***Note: Have selected the best fitness solution for each generation.***	Accuracy with all features: 0.7030199766159058 Accuracy with selected features: 0.7077686190605164 Feature selection improved classification accuracy.

Generation 1 Accuracies: [0.6968839764595032, 0.6552662253379822, 0.7084622979164124, 0.6573471426963806, 0.737648069858551, 0.6016433835029602, 0.656226634979248, 0.7032333612442017, 0.698644757270813, 0.6513712406158447]
Generation 2 Accuracies: [0.6196243762969971, 0.691815197467804, 0.6970974206924438, 0.6433678269386292, 0.6598015427589417, 0.7306050658226013, 0.6630562543869019, 0.5282253623008728, 0.7218012809753418, 0.656226634979248]
Generation 3 Accuracies: [0.698538064956665, 0.5499946475028992, 0.6561732888221741, 0.6655640006065369, 0.7101163268089294, 0.684238612651825, 0.6608152985572815, 0.6113008260726929, 0.6921886801719666, 0.6517447233200073]
Generation 4 Accuracies: [0.7277238368988037, 0.7252694368362427, 0.6237328052520752, 0.6619890928268433, 0.6376587152481079, 0.7362074255943298, 0.6894674897193909, 0.5911855697631836, 0.5866503119468689, 0.696137011051178]
Generation 5 Accuracies: [0.7246825098991394, 0.6768221259117126, 0.7085689902305603, 0.7124106287956238, 0.6956034302711487, 0.6046846508979797, 0.6898409724235535, 0.6309891939163208, 0.6942695379257202, 0.6895742416381836]
Generation 6 Accuracies: [0.5709102749824524, 0.7551488876342773, 0.6927222013473511, 0.6072457432746887, 0.6470494270324707, 0.6448084712028503, 0.6458755731582642, 0.5067228674888611, 0.6054850220680237, 0.6937893629074097]
Generation 7 Accuracies: [0.6745811700820923, 0.6131149530410767, 0.6746878623962402, 0.6870131492614746, 0.6968839764595032, 0.6675381660461426, 0.5576245784759521, 0.7162522673606873, 0.7135310769081116, 0.7040870785713196]
Generation 8 Accuracies: [0.7152385115623474, 0.7021662592887878, 0.7515206336975098, 0.6745811700820923, 0.7161989212036133, 0.5788069367408752, 0.5725643038749695, 0.6912816166877747, 0.6786361932754517, 0.693575918674469]
Generation 9 Accuracies: [0.7223348617553711, 0.5232632756233215, 0.6993383765220642, 0.7333261966705322, 0.7120904922485352, 0.7290577292442322, 0.703766942024231, 0.6499306559562683, 0.7418631911277771, 0.5462063550949097]
Generation 10 Accuracies: [0.6968306303024292, 0.6605485081672668, 0.6900544166564941, 0.6969907283782959, 0.6941628456115723, 0.7074484825134277, 0.7071283459663391, 0.6488101482391357, 0.6549460887908936, 0.6360046863555908]

3. **Analyze the impact of feature selection on model performance and the features selected by the GA**

Impact on Model Performance:

- Feature selection with Genetic Algorithm (GA) can make the model better by choosing the most important features and making the data simpler.
- It helps the model focus on what's essential, which can help it make better predictions on new data.
- However, if important features are left out, it might hurt the model's performance.

Features Selected by the GA:

- The GA picks features that help the model the most in making accurate predictions.
- These features are likely the ones that have a strong connection to what we're trying to predict.
- Understanding which features the GA picks can tell us what parts of the data are crucial for making predictions.
- It's important to check if the selected features really help the model do better.