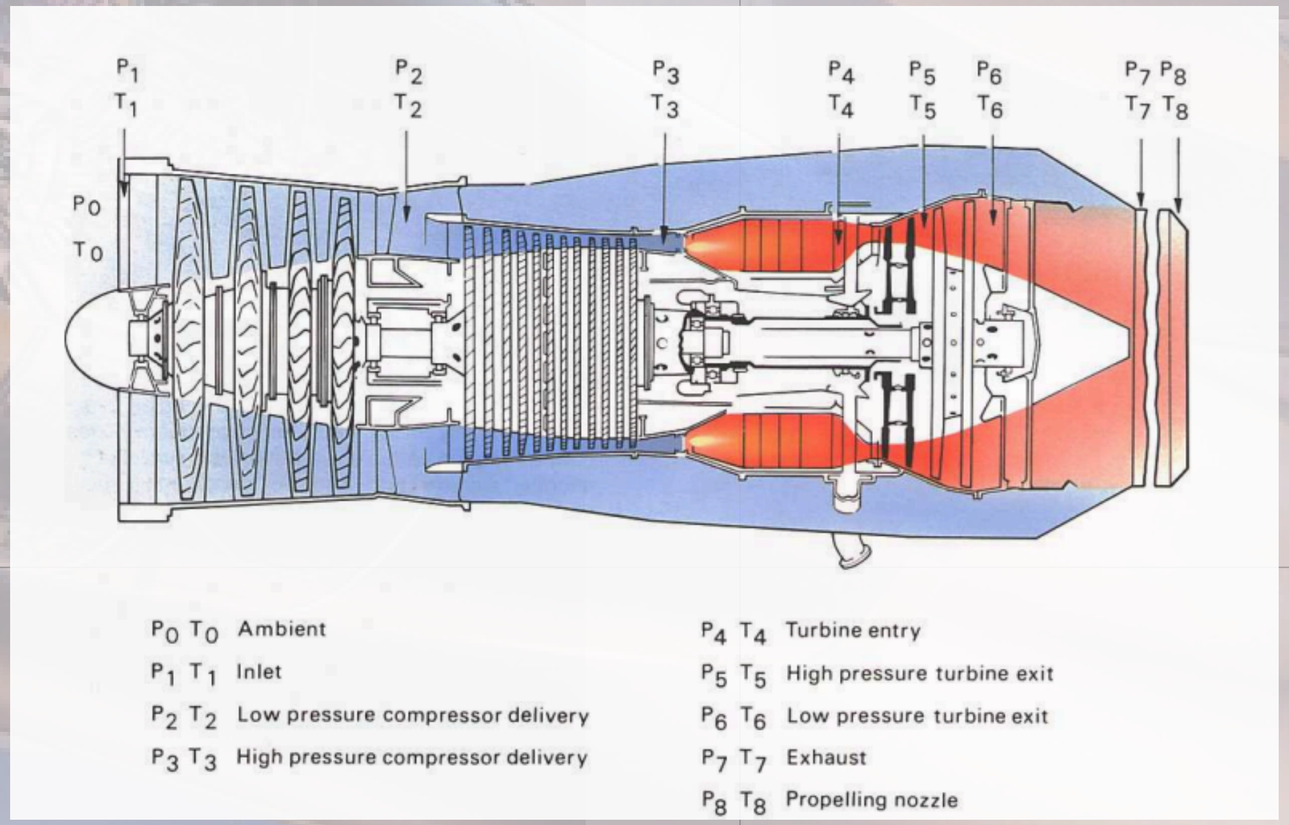
PREDICTIVE MAINTENANCE OF AIRCRAFT ENGINES

PREDICTING REMAINING USEFUL LIFE (RUL) OF TURBOFAN ENGINES WITH LSTM

- AIRCRAFT ENGINES GENERATE LARGE VOLUMES OF TIME-SERIES SENSOR DATA.
- PREDICTING FAILURES IN ADVANCE PREVENTS COSTLY DOWNTIME AND IMPROVES SAFETY.
- GOAL: PREDICT WHETHER AN ENGINE IS CLOSE TO FAILURE (BINARY CLASSIFICATION: FAIL / NO FAIL).



DATASET & PRE-PROCESSING

DATASET: TURBOFAN ENGINE DEGRADATION SIMULATION DATASET FROM THE NASA AMES PROGNOSTICS DATA REPOSITORY (NASA C-MAPSS DATASET).

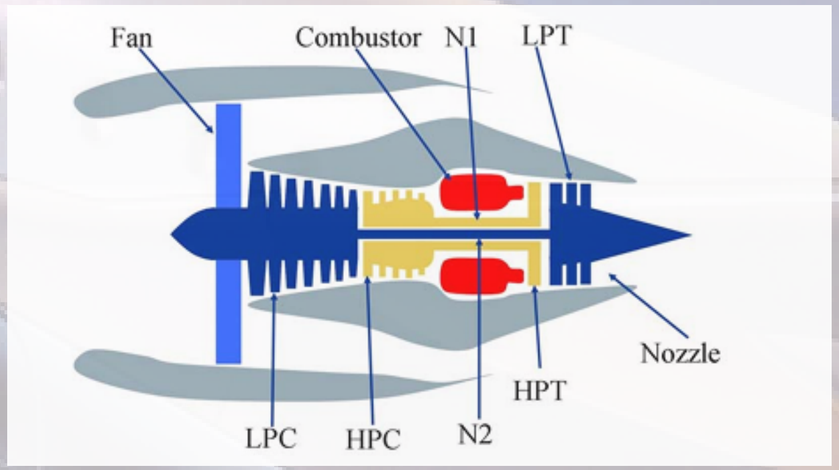
DATA STRUCTURE:

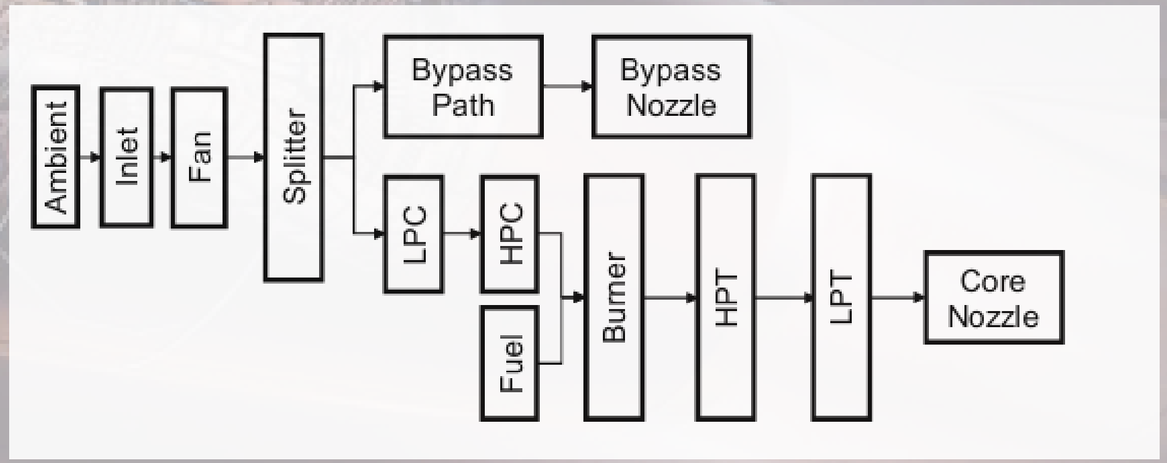
- MULTI-SENSOR TIME-SERIES DATA: THE DATA SIMULATES THE OPERATION OF A FLEET OF ENGINES, EACH RUNNING TO FAILURE.
- FOR EACH ENGINE (UNITNUMBER):
  - CYCLE: THE OPERATIONAL CYCLE (TIME-STEP).
  - OPERATIONAL SETTINGS: 3 SETTINGS (E.G., ALTITUDE, SPEED).
  - SENSOR MEASUREMENTS: 21 SENSORS (E.G., TEMPERATURE, PRESSURE, FAN SPEED).
- THE "GROUND TRUTH": THE DATASET PROVIDES THE TRUE RUL FOR THE TEST SET, ALLOWING US TO VALIDATE OUR MODEL. FOR THE TRAINING SET, RUL IS CALCULATED AS (MAX CYCLES - CURRENT CYCLE).

NASA C-MAPSS-1 Turbofan Engine Degradation Dataset

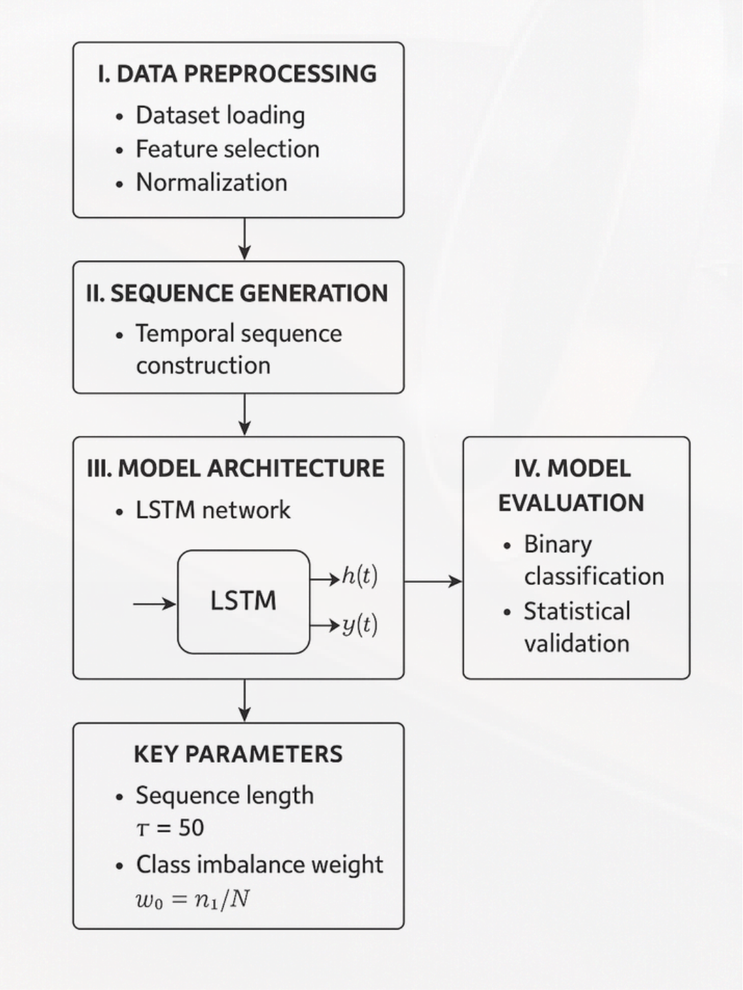
Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation

k kaggle.com





WORKFLOW & PRE-PROCESSING



DATA ACQUISITION:
LOADING THE NASA C-MAPSS DATASET.

DATA PRE-PROCESSING & VISUALIZATION:
FEATURE ENGINEERING: CALCULATING THE RUL COLUMN FOR THE TRAINING DATA BY FINDING THE MAX CYCLE FOR EACH ENGINE AND SUBTRACTING THE CURRENT CYCLE.
VISUALIZATION (EDA): PLOTTING SENSOR DATA OVER TIME TO IDENTIFY TRENDS. EARLY CYCLES SHOW STABLE SENSOR READINGS, WHILE READINGS TREND UP OR DOWN AS THE ENGINE APPROACHES FAILURE.
FEATURE SELECTION: REMOVING STATIC FEATURES (SENSORS OR SETTINGS THAT DON'T CHANGE).
NORMALIZATION: SCALING ALL SENSOR AND OPERATIONAL DATA TO A COMMON RANGE (E.G., -1 TO 1) USING MINMAXSCALER.

DATA PREPARATION (FOR LSTM):
WINDOWING: CONVERTING THE TIME-SERIES DATA INTO "SEQUENCES". USING A WINDOW OF 50 CYCLES TO PREDICT THE RUL AT THE 50TH CYCLE. THE MODEL LEARNS FROM THE PATTERN OF THE LAST 50 CYCLES.
MODEL TRAINING:
BUILDING AND TRAIN THE LSTM MODEL ON THE SEQUENCED, NORMALIZED TRAINING DATA.
MODEL EVALUATION:
USING THE TRAINED MODEL TO PREDICT RUL ON THE UNSEEN TEST DATA AND COMPARE ITS PREDICTIONS TO THE TRUE RUL VALUES.

RAW DATA → CLEANED & NORMALIZED → TIME-WINDOWED SEQUENCES → LSTM TRAINING → RUL PREDICTION & EVALUATION

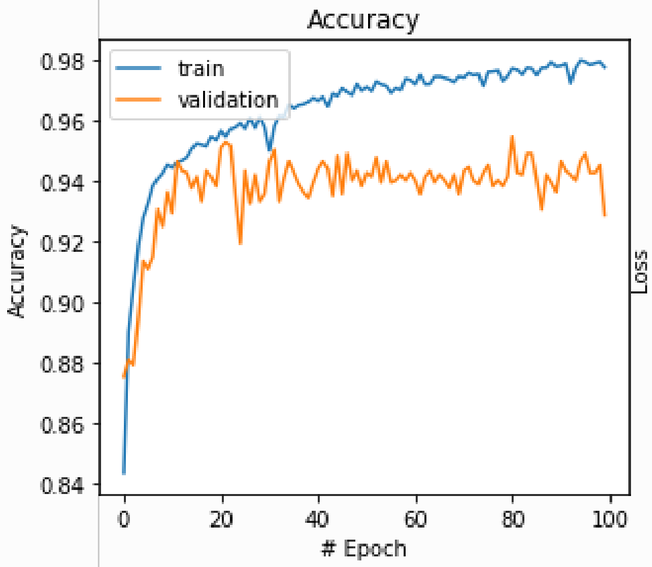
EVALUATION & CONCLUSION

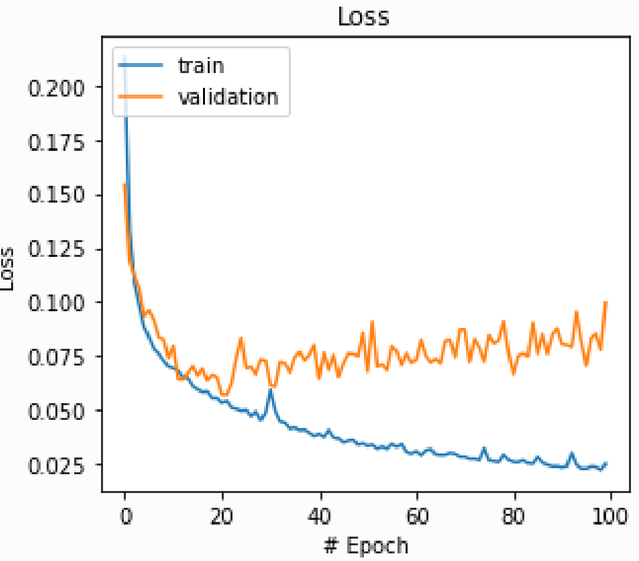
MODEL 1: ONE HIDDEN LAYER

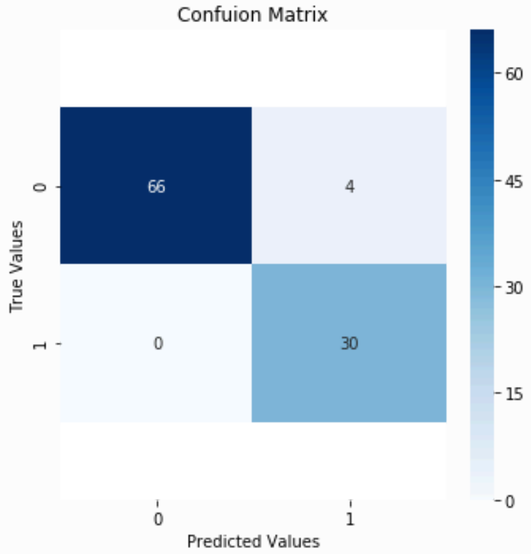
F1 Score: 0.9375

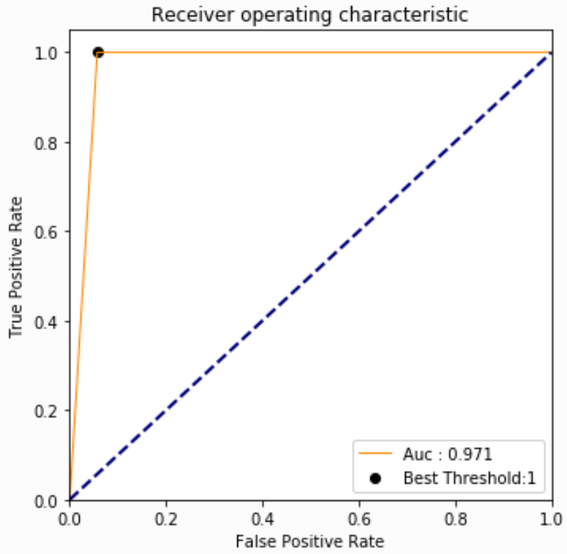
	precision	recall	f1-score	support
0	1.00	0.94	0.97	70
1	0.88	1.00	0.94	30
accuracy			0.96	100
macro avg	0.94	0.97	0.95	100
weighted avg	0.96	0.96	0.96	100

Area under curve : 0.9714285714285714









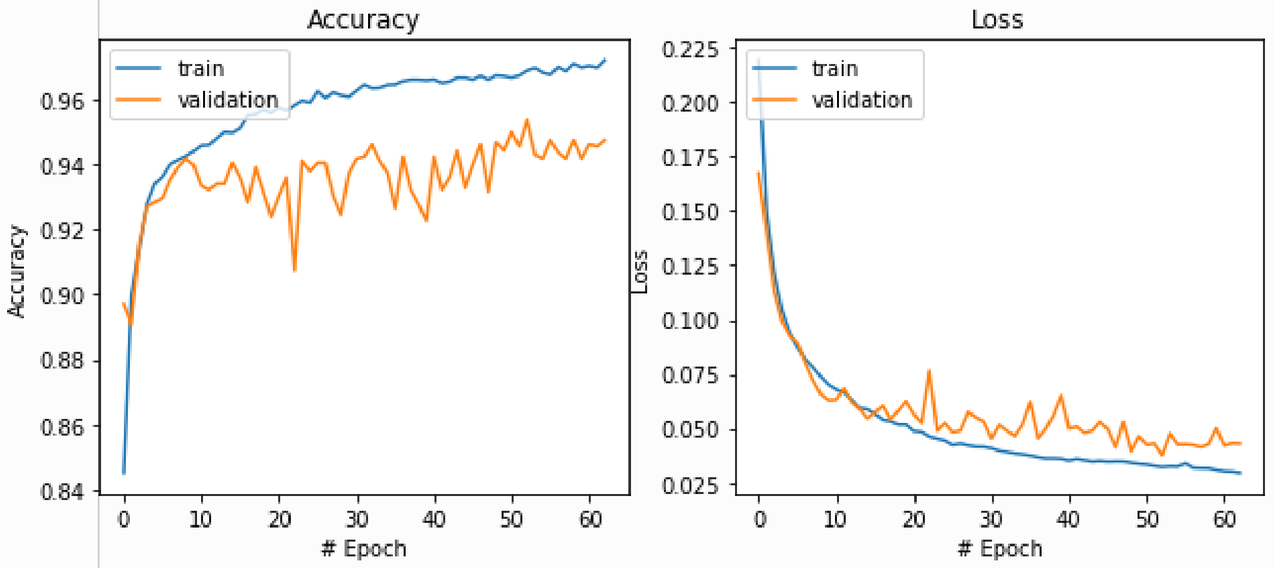
EVALUATION & CONCLUSION

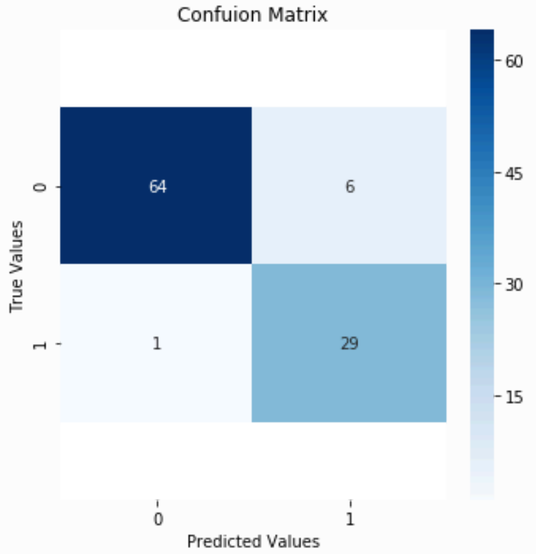
MODEL 2: ONE HIDDEN LAYER WITH EARLY STOPPING

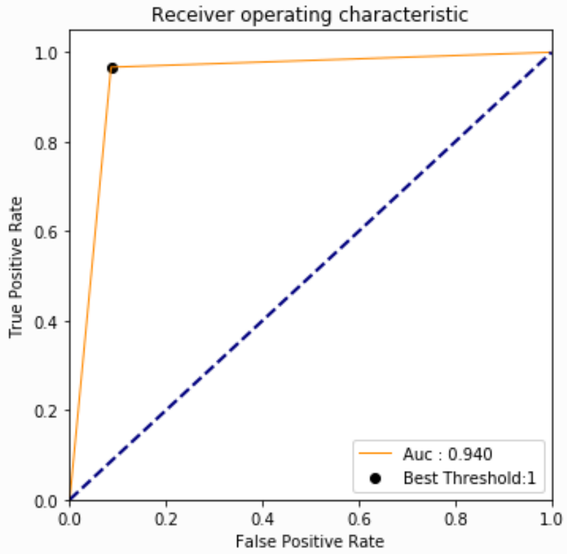
F1 Score: 0.8923076923076922

	precision	recall	f1-score	support
0	0.98	0.91	0.95	70
1	0.83	0.97	0.89	30
accuracy			0.93	100
macro avg	0.91	0.94	0.92	100
weighted avg	0.94	0.93	0.93	100

Area under curve : 0.9404761904761906







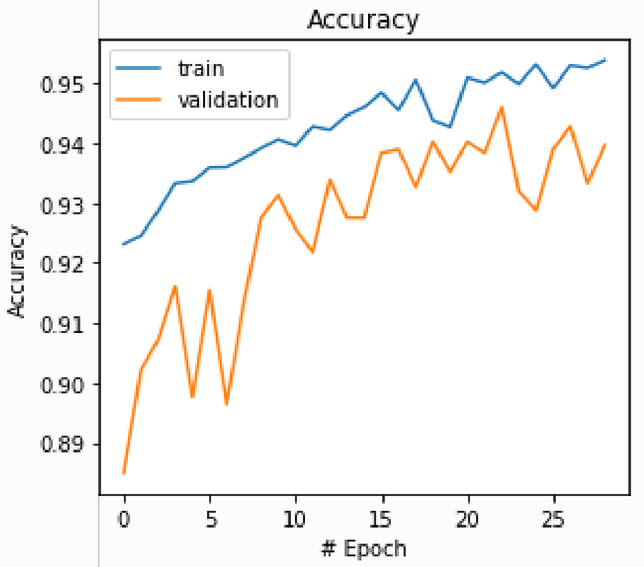
EVALUATION & CONCLUSION

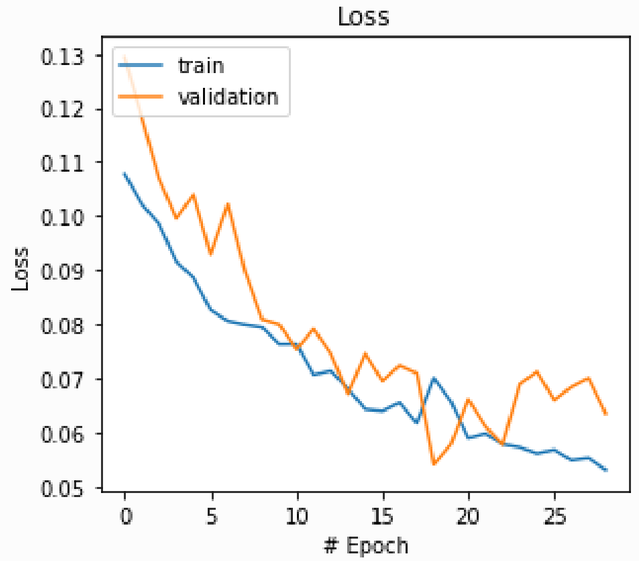
MODEL 3: ONE HIDDEN LAYER WITH EARLY STOPPING AND DROPOUT

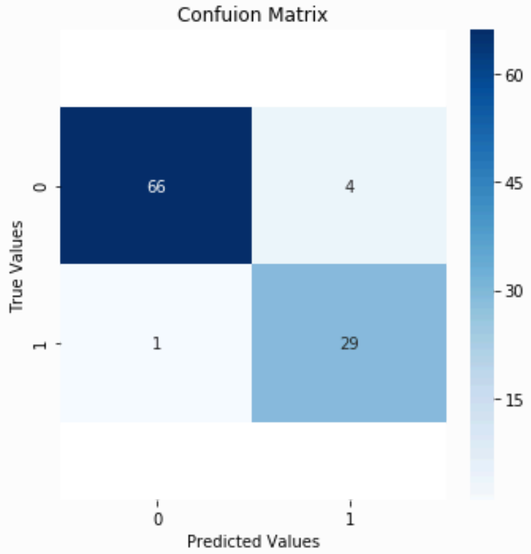
F1 Score: 0.9206349206349207

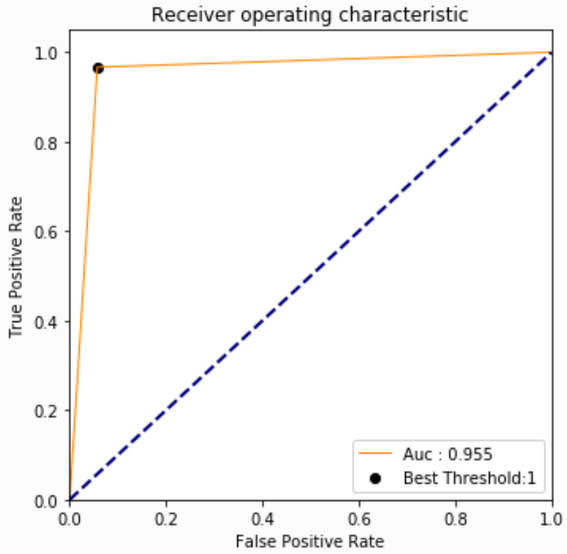
	precision	recall	f1-score	support
0	0.99	0.94	0.96	70
1	0.88	0.97	0.92	30
accuracy			0.95	100
macro avg	0.93	0.95	0.94	100
weighted avg	0.95	0.95	0.95	100

Area under curve : 0.9547619047619048









EVALUATION & CONCLUSION

MODEL 4: TWO HIDDEN LAYER

F1 Score: 0.923076923076923

	precision	recall	f1-score	support
0	1.00	0.93	0.96	70
1	0.86	1.00	0.92	30
accuracy			0.95	100
macro avg	0.93	0.96	0.94	100
weighted avg	0.96	0.95	0.95	100

Area under curve : 0.9642857142857143