

MỘT SỐ PHƯƠNG PHÁP SINH DỮ LIỆU CHO BÀI TOÁN PHÂN LỚP CHUỖI THỜI GIAN

Nhóm 3

Viện Toán Ứng dụng và Tin học
Đại học Bách khoa Hà Nội

Ngày 9 tháng 2 năm 2022

Danh sách các thành viên:

- 1) Nguyễn Hữu Thuật - 20185410
- 2) Trang Hải Long - 20185382
- 3) Lại Tiến Long - 20185376

Nội dung

- 1 Giới thiệu đề tài
 - Lý do chọn đề tài
 - Thực trạng
- 2 Lý thuyết cơ sở
 - Chuỗi thời gian
 - Máy học
 - Một số mô hình
- 3 Các phương pháp sinh dữ liệu
 - Jittering
 - Rotation
 - Scaling
 - Magnitude Warping
 - Permutation
 - Window Slice
 - Time Warping
 - Window Warping
 - SuboPtimAI Warped time series geNErator (SPAWNER)
 - Weighted DTW Barycentric Averaging (wDBA)
 - Random Guided Warping (RGW)
 - Discriminative Guided Warping (DGW)
- 4 Ứng dụng
 - Dữ liệu
 - Phương pháp thực nghiệm
- 5 Kết quả và kết luận

Lý do chọn đề tài

- Dữ liệu về chuỗi thời gian là rất phổ biến trong đời sống hiện nay.

Lý do chọn đề tài

- Dữ liệu về chuỗi thời gian là rất phổ biến trong đời sống hiện nay.
- Nhưng dữ liệu được phân lớp, gán nhãn thì còn hạn chế do chi phí của việc gán nhãn dữ liệu là không hề nhỏ

- Các kỹ thuật sinh dữ liệu đã được tìm thấy, hữu ích trong các lĩnh vực như NLP- xử lý ngôn ngữ tự nhiên và Computer Vision-thị giác máy tính.

- **Người ta có đang nghiên cứu về Sinh dữ liệu nhiều hay không ?**

- Người ta có đang nghiên cứu về Sinh dữ liệu nhiều hay không ?
- Nghiên cứu thì nghiên cứu bằng những phương pháp nào ?

- **Người ta có đang nghiên cứu về Sinh dữ liệu nhiều hay không ?**
 - **Nghiên cứu thì nghiên cứu bằng những phương pháp nào ?**
- + Một số phương pháp thêm dữ liệu:
- Collect more data
 - Data synthesis
 - Data Augmentation

Nội dung

- 1 Giới thiệu đề tài
 - Lý do chọn đề tài
 - Thực trạng
- 2 Lý thuyết cơ sở
 - Chuỗi thời gian
 - Máy học
 - Một số mô hình
- 3 Các phương pháp sinh dữ liệu
 - Jittering
 - Rotation
 - Scaling
 - Magnitude Warping
 - Permutation
 - Window Slice
 - Time Warping
 - Window Warping
 - SuboPtimAI Warped time series geNErator (SPAWNER)
 - Weighted DTW Barycentric Averaging (wDBA)
 - Random Guided Warping (RGW)
 - Discriminative Guided Warping (DGW)
- 4 Ứng dụng
 - Dữ liệu
 - Phương pháp thực nghiệm
- 5 Kết quả và kết luận

Chuỗi thời gian

Chuỗi thời gian là 1 tập quan sát X_t , mỗi quan sát được ghi lại tại 1 thời điểm cụ thể

White Noise

Chuỗi X_t là chuỗi các biến ngẫu nhiên không tương quan là 1 chuỗi nhiễu trắng với kỳ vọng là 0 và phương sai là σ^2
Ký hiệu: $X_t \sim WN(0, \sigma^2)$

Thành phần của 1 chuỗi thời gian

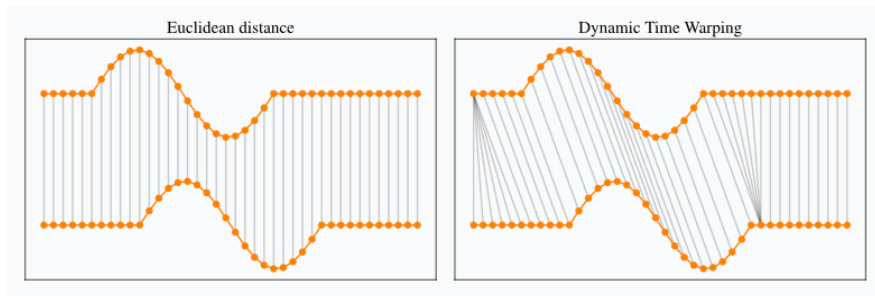
Bao gồm 4 thành phần:

- + Thành phần xu hướng (trend)
- + Thành phần mùa (seasonal)
- + Thành phần dư (residual)
- + Thành phần chu kỳ (cycle)

Tính dừng

- Dừng chặt: X_t là chuỗi dừng chặt nếu: tồn tại thời gian dịch chuyển h sao cho: X_{t+h} có cùng phân phối với X_t
- Dừng yếu: X_t là chuỗi dừng yếu nếu:
 - + $\mu_X(r)$ độc lập với t
 - + $\gamma_X(t+h, t)$ độc lập với t với mọi h

- **Định nghĩa:** Dynamic Time Warping là cách để so sánh hai chuỗi thời gian không đồng bộ với nhau.



Hình: Khoảng cách Euclid và khoảng cách DTW

- Máy học (Machine Learning)

Khái niệm

- Máy học (Machine Learning)
- Học sâu (Deep Learning)

Khái niệm

- Máy học (Machine Learning)
- Học sâu (Deep Learning)
- Độ chính xác (Accuracy)

Các mô hình đề xuất sử dụng

- + Logistic Regression
- + Linear Discriminant Analysis
- + Gaussian Naive Bayes
- + Decision Tree Classifier
- + Random Forest Classifier
- + K-Nearest Neighbor
- + Ada Boost
- + XG Boost
- + Light GBM

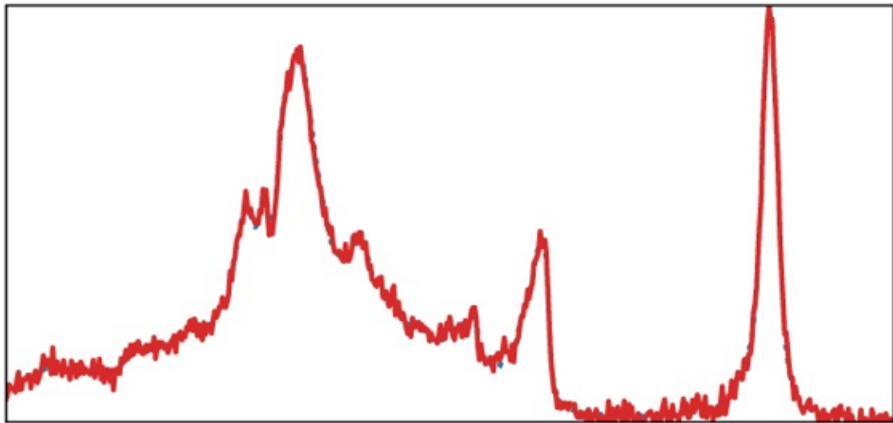
Các mô hình đề xuất sử dụng

- + VGG
- + LSTM-1
- + LSTM
- + LSTM-2
- + BLSTM
- + BLSTM-2
- + LSTM-FCN
- + Res-Net
- + MLP
- + Le-Net

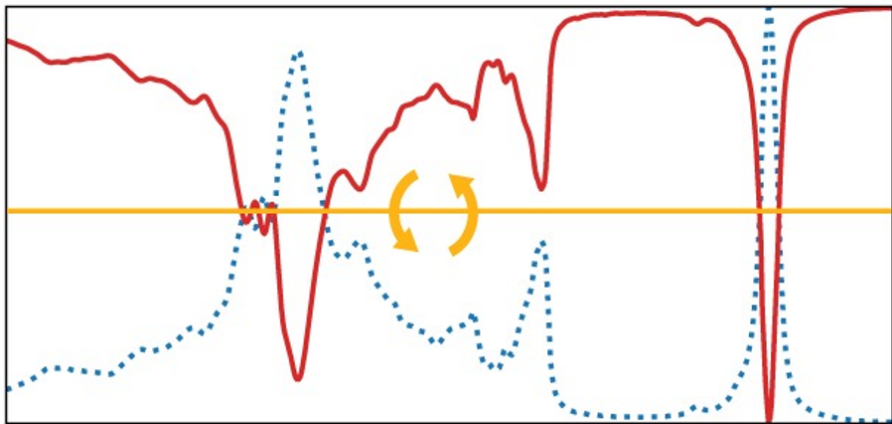
Nội dung

- 1 Giới thiệu đề tài
 - Lý do chọn đề tài
 - Thực trạng
- 2 Lý thuyết cơ sở
 - Chuỗi thời gian
 - Máy học
 - Một số mô hình
- 3 Các phương pháp sinh dữ liệu
 - Jittering
 - Rotation
 - Scaling
 - Magnitude Warping
 - Permutation
 - Window Slice
 - Time Warping
 - Window Warping
 - SuboPtimAI Warped time series geNErator (SPAWNER)
 - Weighted DTW Barycentric Averaging (wDBA)
 - Random Guided Warping (RGW)
 - Discriminative Guided Warping (DGW)
- 4 Ứng dụng
 - Dữ liệu
 - Phương pháp thực nghiệm
- 5 Kết quả và kết luận

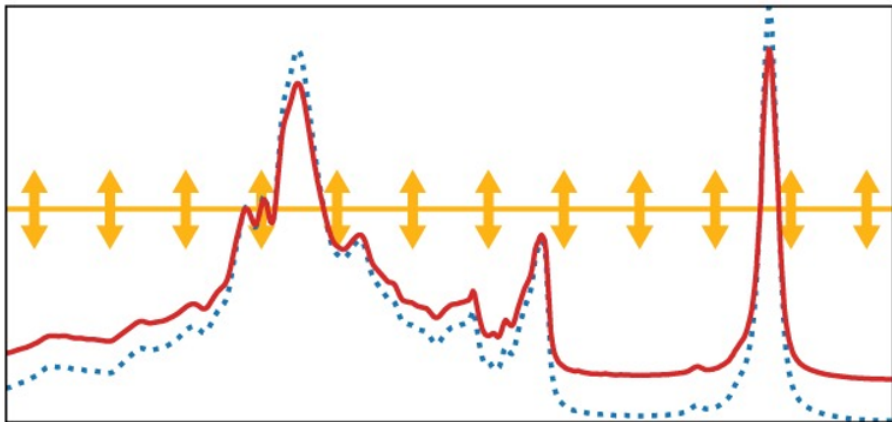
Jittering



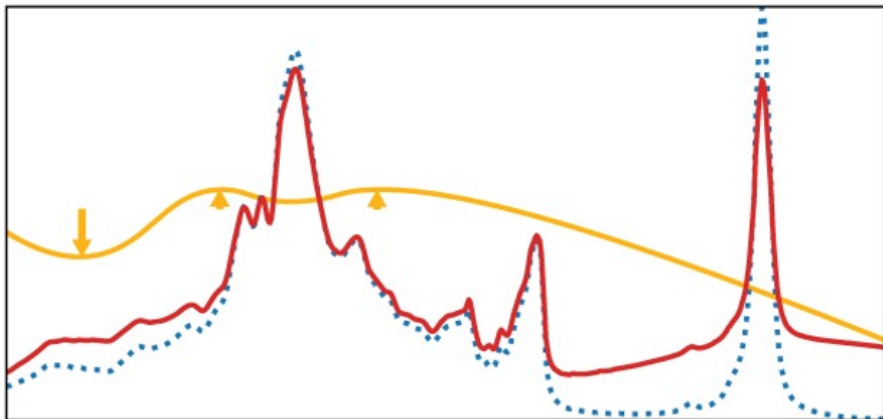
Rotation



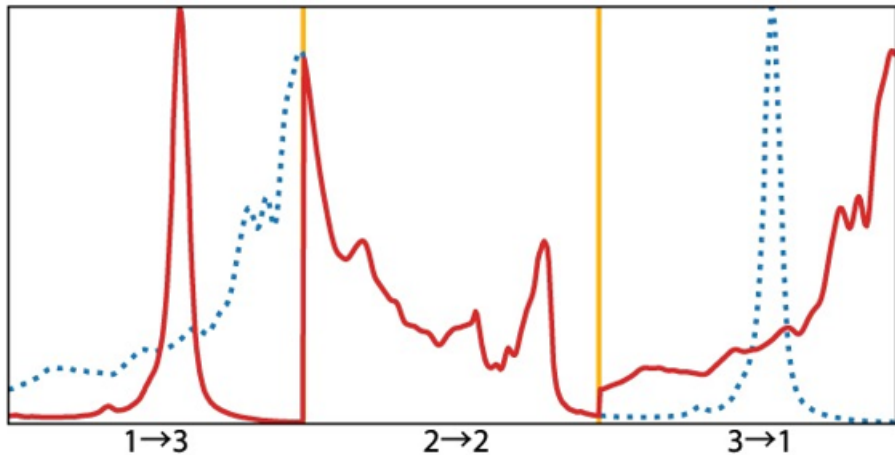
Scaling



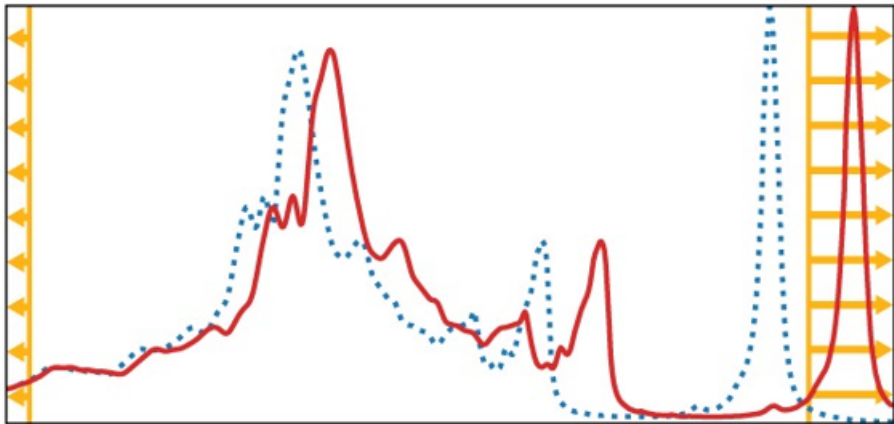
Magnitude Warping



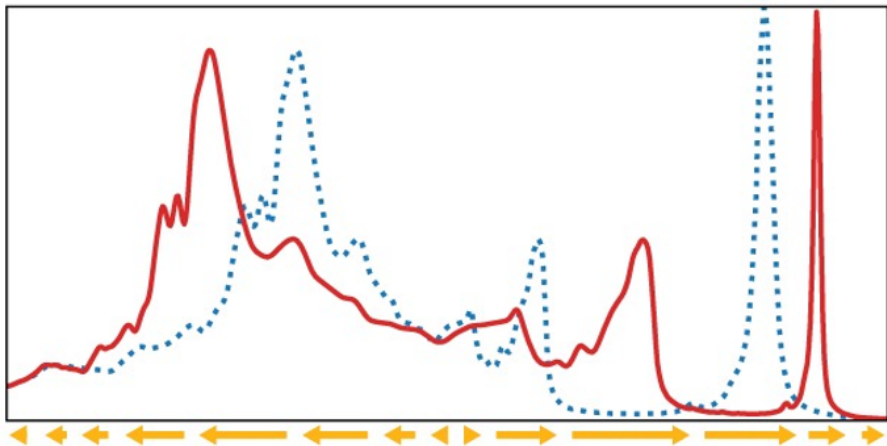
Permutation



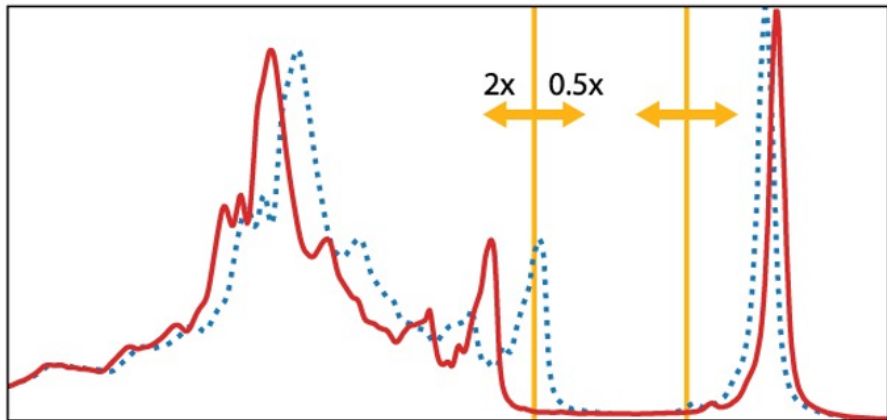
Window Slice



Time Warping



Window Warping

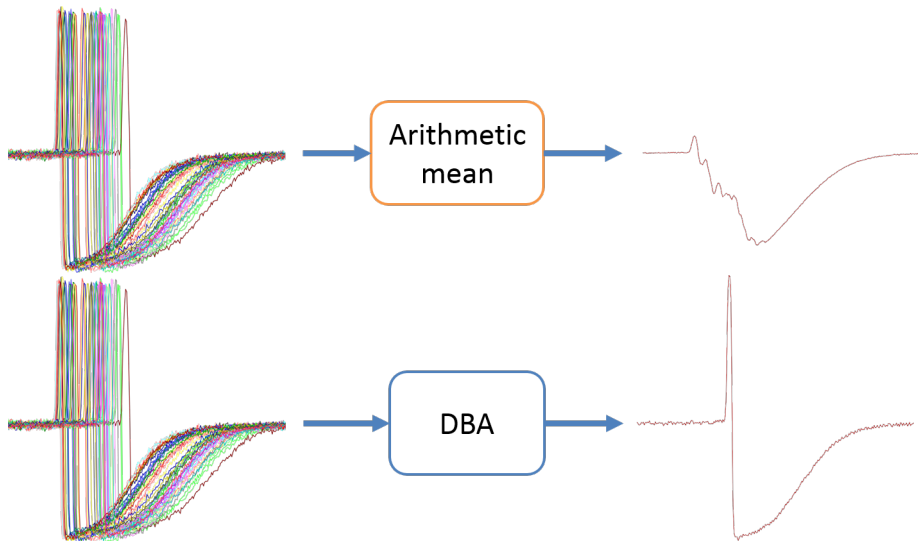


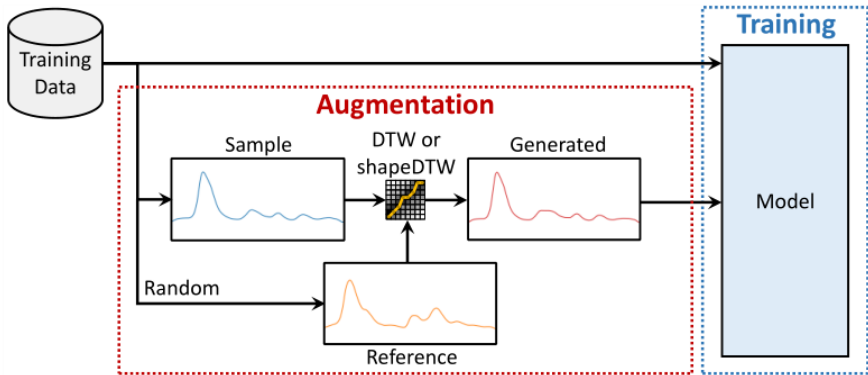
- Sinh một chuỗi thời gian mới từ hai chuỗi thời gian X_1 và X_2 cùng nhãn.
- Sử dụng DTW với hai chuỗi thời gian thu được đường căn chỉnh:

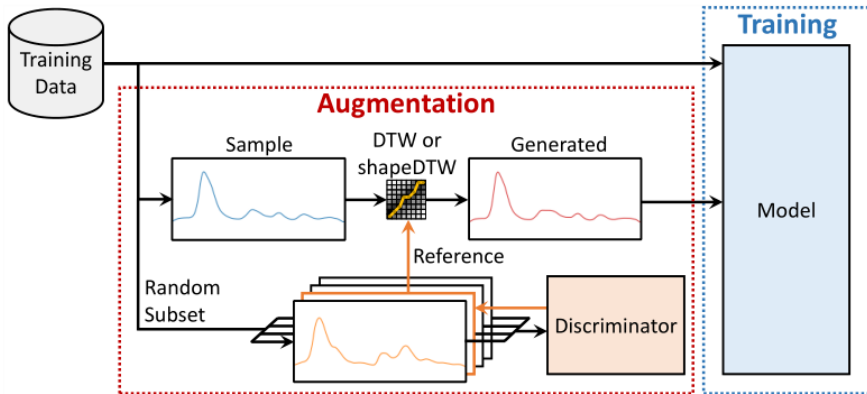
$$W = [w_1, w_2, \dots, w_P].$$

- Hợp nhất hai chuỗi thời gian trên bằng cách lấy giá trị trung bình các phần tử tương ứng của hai chuỗi $\rightarrow X^*$
- X^* cũng có thể lấy sao cho $x^* \in X^*$:

$$x^* \sim N(\mu, \sigma^2), \mu = 0.5(x_1^* + x_2^*), \sigma = 0.05\|x_1^* - x_2^*\|$$







Nội dung

- 1 Giới thiệu đề tài
 - Lý do chọn đề tài
 - Thực trạng
- 2 Lý thuyết cơ sở
 - Chuỗi thời gian
 - Máy học
 - Một số mô hình
- 3 Các phương pháp sinh dữ liệu
 - Jittering
 - Rotation
 - Scaling
 - Magnitude Warping
 - Permutation
 - Window Slice
 - Time Warping
 - Window Warping
 - SuboPtimAI Warped time series geNErator (SPAWNER)
 - Weighted DTW Barycentric Averaging (wDBA)
 - Random Guided Warping (RGW)
 - Discriminative Guided Warping (DGW)
- 4 Ứng dụng
 - Dữ liệu
 - Phương pháp thực nghiệm
- 5 Kết quả và kết luận

- Bộ *CBF* được trích từ tập *UCR Suite*

- Bộ *CBF* được trích từ tập *UCR Suite*
- Đặc điểm bộ dữ liệu:
 - + 930 bản ghi

- Bộ *CBF* được trích từ tập *UCR Suite*
- Đặc điểm bộ dữ liệu:
 - + 930 bản ghi
 - + Mỗi bản ghi là một chuỗi thời gian

- Bộ *CBF* được trích từ tập *UCR Suite*
- Đặc điểm bộ dữ liệu:
 - + 930 bản ghi
 - + Mỗi bản ghi là một chuỗi thời gian
 - + Số nhãn: 3

- Bộ *CBF* được trích từ tập *UCR Suite*
- Đặc điểm bộ dữ liệu:
 - + 930 bản ghi
 - + Mỗi bản ghi là một chuỗi thời gian
 - + Số nhãn: 3
 - + Số trường: 128

Phương pháp thực nghiệm

- Chia bộ dữ liệu thành 2 tập: 30:900

Phương pháp thực nghiệm

- Chia bộ dữ liệu thành 2 tập: 30:900
- Sinh dữ liệu dựa với tập huấn luyện

Phương pháp thực nghiệm

- Chia bộ dữ liệu thành 2 tập: 30:900
- Sinh dữ liệu dựa với tập huấn luyện
- Thực nghiệm chất lượng của các tập dữ liệu mới sinh

Phương pháp thực nghiệm

- Chia bộ dữ liệu thành 2 tập: 30:900
- Sinh dữ liệu dựa với tập huấn luyện
- Thực nghiệm chất lượng của các tập dữ liệu mới sinh
- Kiểm định và kết luận

Nội dung

- 1 Giới thiệu đề tài
 - Lý do chọn đề tài
 - Thực trạng
- 2 Lý thuyết cơ sở
 - Chuỗi thời gian
 - Máy học
 - Một số mô hình
- 3 Các phương pháp sinh dữ liệu
 - Jittering
 - Rotation
 - Scaling
 - Magnitude Warping
 - Permutation
 - Window Slice
 - Time Warping
 - Window Warping
 - SuboPtimAI Warped time series geNErator (SPAWNER)
 - Weighted DTW Barycentric Averaging (wDBA)
 - Random Guided Warping (RGW)
 - Discriminative Guided Warping (DGW)
- 4 Ứng dụng
 - Dữ liệu
 - Phương pháp thực nghiệm
- 5 Kết quả và kết luận

Kết quả

	None	Jittering	Scaling	Rotation	Permutation	Magnitude Warping
Logistic Regression	0.85	0.85	0.84	0.46	0.34	0.86
Linear Discriminant Analysis	0.84	0.87	0.79	0.96	0.36	0.41
Gaussian Naive Bayes	0.89	0.90	0.90	0.58	0.86	0.91
Decision Tree Classifier	0.76	0.80	0.72	0.47	0.54	0.73
Random Forest Classifier	0.87	0.93	0.93	0.78	0.84	0.95
KNN	0.77	0.85	0.85	0.85	0.84	0.85
Ada Boost	0.81	0.81	0.73	0.49	0.49	0.79
XGBoost	0.77	0.82	0.85	0.54	0.73	0.8

Kết quả

	Time Warping	Slicing	Window Warping	SPAWNER	wDBA	RGW	DGW
Logistic Regre- ssion	0.75	0.95	0.89	0.84	0.83	0.89	0.91
Linear Discrim- inant Analysis	0.78	0.94	0.78	0.84	0.84	0.78	0.89
Gaussian Naive Bayes	0.82	0.91	0.88	0.90	0.78	0.90	0.90
Decision Tree Classifier	0.73	0.83	0.83	0.78	0.72	0.77	0.79
Random Forest Classifier	0.95	0.99	0.95	0.95	0.91	0.97	0.99
KNN	1.0	0.97	0.97	0.91	0.84	0.93	0.98
Ada Boost	0.66	0.81	0.82	0.90	0.70	0.80	0.89
XGBoost	0.75	0.94	0.88	0.94	0.78	0.88	0.95

Kết quả

	None	Jittering	Scaling	Rotation	Permutation	Magnitude Warping
VGG	0.41	1.0	1.0	0.84	0.98	0.99
LSTM-1	0.33	0.43	0.53	0.33	0.33	0.44
LSTM	0.51	0.33	0.53	0.52	0.33	0.55
LSTM-2	0.33	0.33	0.52	0.51	0.44	0.40
BLSTM	0.33	0.33	0.35	0.52	0.34	0.44
BLSTM-2	0.42	0.48	0.76	0.81	0.77	0.94
LSTM-FCN	0.33	0.36	0.61	0.95	0.70	0.57
Res-Net	0.35	0.33	0.33	0.33	0.33	0.33
MLP	0.5	0.90	0.92	0.87	0.88	0.89
Le-Net	0.36	0.89	0.90	0.84	0.88	0.90

Kết quả

	Time Warping	Slicing	Window Warping	SPAWNER	wDBA	RGW	DGW
VGG	0.99	0.99	0.98	0.96	0.97	0.98	0.97
LSTM-1	0.61	0.69	0.47	0.47	0.48	0.55	0.53
LSTM	0.57	0.59	0.78	0.82	0.36	0.48	0.51
LSTM-2	0.55	0.62	0.67	0.75	0.82	0.76	0.87
BLSTM	0.51	0.49	0.72	0.70	0.77	0.75	0.81
BLSTM-2	0.97	0.93	0.97	0.99	0.98	0.97	0.96
LSTM-FCN	0.95	0.97	0.95	0.60	0.99	0.95	0.91
Res-Net	0.33	0.38	0.61	0.59	0.98	0.99	0.96
MLP	0.89	0.96	0.95	0.95	0.93	0.95	0.95
Le-Net	0.91	0.95	0.95	0.97	0.97	0.98	0.97

- Không khẳng định Phương pháp nào là tốt nhất

Kết luận

- Không khẳng định Phương pháp nào là tốt nhất
- Hiệu quả của Phương pháp sinh dữ liệu

Kết luận

- Không khẳng định Phương pháp nào là tốt nhất
- Hiệu quả của Phương pháp sinh dữ liệu
- Vai trò của dữ liệu đối với các mô hình học sâu.



THANK YOU
FOR WATCHING