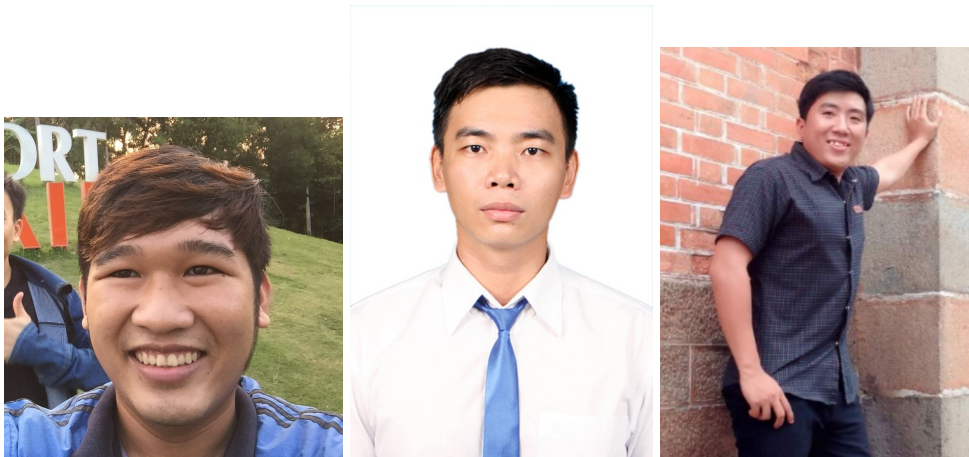



<b>Họ và tên (IN HOA)</b>	NGUYỄN HỮU VINH (CH2001042) NGUYỄN PHÚC THỊNH (CH2001038) NGUYỄN XUÂN THẢO (CH2004016)
<b>Ảnh</b>	
<b>Số buổi vắng</b>	0
<b>Bonus</b>	19
<b>Tên đề tài (VN)</b>	DỰ ĐOÁN TỐC ĐỘ CHẠY VÀ CÂN NẶNG QUA DỮ LIỆU CUỘC SỐNG HẰNG NGÀY
<b>Tên đề tài (EN)</b>	
<b>Giới thiệu</b>	<ul style="list-style-type: none"> <li>● Vấn đề cần giải quyết : dự đoán được cân nặng và tốc độ chạy dựa trên các chỉ số cơ thể</li> <li>● Lý do chọn đề tài: sức khỏe là vấn đề rất quan trọng trong đời sống con người hiện đại. Đo lường và dự báo được các chỉ số cuộc sống có thể giúp dự báo trước được bệnh tật và khuyến nghị điều chỉnh lối sống.</li> <li>● Input-output:</li> </ul>

	<ul style="list-style-type: none"><li>- Input: Các chỉ số ghi nhận hàng ngày: nhịp tim, nhịp thở, số bước chạy, chế độ ăn,...</li><li>- Output: cân nặng và tốc độ chạy.</li></ul>  <pre>graph LR; A["Lifelog Sức khỏe Nhịp tim Chiều cao Cân nặng Giấc ngủ"] -- "Làm sạch dữ liệu" --&gt; B["Dữ liệu đầu vào"]; B -- "Linear Regression, Random Forest Regression" --&gt; C["Cân nặng, tốc độ chạy"]</pre>
<b>Mục tiêu</b>	<ul style="list-style-type: none"><li>● Nắm bắt được các yếu tố chính trong lifelog ảnh hưởng lớn đến sức khỏe con người</li><li>● Tìm ra mô hình thích hợp để dự đoán tương quan giữa các yếu tố trong lifelog và cân nặng, tốc độ chạy. Đánh giá độ chính xác của mô hình đó.</li><li>● Phục vụ cho các ứng dụng khác: camera giám sát, giải trí, chăm sóc sức khỏe,...</li></ul>
<b>Nội dung và phương pháp thực hiện</b>	<ul style="list-style-type: none"><li>● Nội dung nghiên cứu<ul style="list-style-type: none"><li>- Đọc và hiểu rõ các tài liệu trong phần tham khảo.</li><li>- Tập hợp dữ liệu liên quan từ các tài liệu tham khảo.</li><li>- Đánh giá lại các phương pháp trong phần tài liệu tham khảo.</li><li>- Sử dụng bộ dữ liệu của ImageCLEF Lifelog 2020-SPLL để dự đoán sự thay đổi của tốc độ chạy (speed) và cân nặng kể từ đầu kỳ báo cáo đến cuối kỳ báo cáo, với cân nặng tính bằng kilogam và tốc độ dựa trên sự thay đổi số giây được sử dụng trên mỗi kilometer.</li></ul></li><li>● Phương pháp thực hiện:<ul style="list-style-type: none"><li>- Làm mịn và khử nhiễu có thể loại bỏ các đột biến, xu hướng và ngoại lệ không mong muốn ra khỏi mô hình.</li><li>- Đồng bộ hóa các khoảng thời gian khác nhau thành một khoảng thời gian để tiện cho việc xử lý dữ liệu sau này. Ví dụ, tính tổng lượng calo đốt cháy cho mỗi hoạt động và tổng lượng calo tiêu thụ mỗi ngày.</li></ul></li></ul>

- Xử lý các giá trị bị thiếu bằng cách điền vào với giá trị phía trước nó. Đặc biệt với các thuộc tính không phải giá trị, giá trị trung bình sẽ được điền vào. Các điểm ngoại lệ cũng được xóa nhằm giảm tác động của chúng đến kết quả cuối cùng.
- Xác định độ tương quan giữa các biến trong bộ dữ liệu, để chọn ra các biến có độ tương quan cao nhất với biến cần dự đoán và tất nhiên là các biến đó phải có cùng miền (domain) với biến cần dự đoán.
- Dùng mô hình Linear Regression, Random Forest Regression dự đoán 'weight' (cân nặng) và 'speed' (tốc độ chạy). Lần lượt thêm các biến độc lập có độ tương quan giảm dần theo khoảng để dạy (train) cho mô hình, cho đến khi nào các chỉ số Mean Squared Error (MSE), Mean Absolute Error (MAE) và Root Mean Squared Error (RMSE) giảm đến mức nhỏ nhất có thể.
- Xây dựng mô hình nhận thức nhiều lớp (multilayer perception model) bằng cách sử dụng ReLU activation function. Sau đó, xây dựng 3 mô hình CNN, kế tiếp là 3 mô hình LSTM và cuối cùng là mô hình GRU.

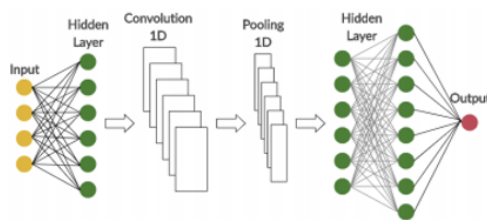


Fig. 1: First CNN model architecture

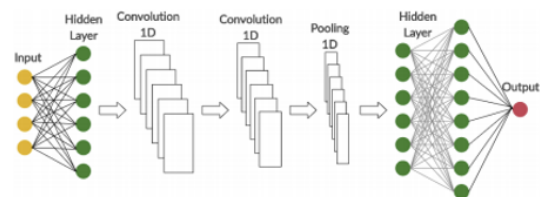


Fig. 2: Second CNN model architecture

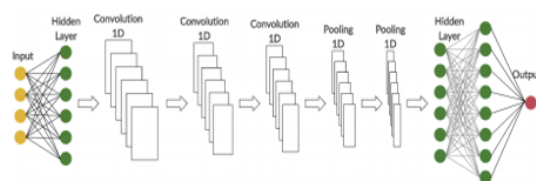


Fig. 3: Third CNN model architecture

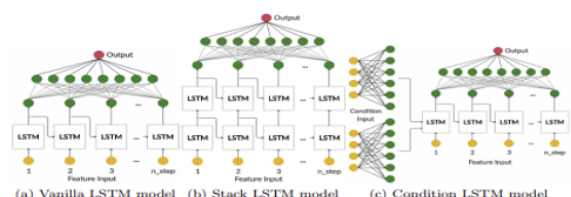


Fig. 4: LSTM-like model architectures

## Kết quả dự kiến

- So sánh MAE, MSE, RMSE theo bảng của các mô hình khi chạy trên dữ liệu từng người và bộ dữ liệu tất cả các người trong tập dữ liệu theo từng hạng mục dự đoán (speed, weight).

Time Prediction Model	Validation MSE	Validation MAE	Train MSE	Train MAE
VanillaLSTMmodel_one_attribute_time-steps_5	3872.775	43.5962	3593.101	43.40623
VanillaLSTMmodel_one_attribute_time-steps_7	3922.615	47.3889	3554.712	42.87104
Condition_LSTM_one_attribute_time-steps_5	4023.349	43.10283	4874.171	51.20744
StackLSTMmodel_one_attribute_time-steps_5	4039.655	44.82555	4191.725	45.65004
StackLSTMmodel_one_attribute_time-steps_3	4044.094	46.14255	2586.746	35.38247
Second_CNN_model_one_attribute_time-steps_5	4046.895	44.3603	4435.651	48.17548
GRU_model_one_attribute_time-steps_7	4076.255	47.21735	4922.799	52.2997
Condition_LSTM_one_attribute_time-steps_7	4216.521	48.09542	4940	52.39553
First_CNN_model_one_attribute_time-steps_7	4263.523	48.82619	2918.934	39.79049
MLPmodel_one_attribute_time-steps_7	4265.8	46.76684	4693.186	50.9879

Weight Prediction Model	Validation MSE	Validation MAE	Train MSE	Train MAE
Condition_LSTM_weight_one_attribute_time-steps_14	0.204387	0.320392	0.350457	0.260382
LR_model_weight_one_attribute_time-steps_7	0.211679	0.318684	0.426128	0.305971
VanillaLSTMmodel_weight_one_attribute_time-steps_7	0.211884	0.326093	0.379312	0.285081
LR_model_weight_one_attribute_time-steps_14	0.215518	0.33098	0.448237	0.296486
Condition_LSTM_weight_one_attribute_time-steps_7	0.216421	0.318217	0.353498	0.25426
VanillaLSTMmodel_weight_one_attribute_time-steps_14	0.217808	0.341755	0.402532	0.285623
StackLSTMmodel_weight_one_attribute_time-steps_7	0.218527	0.337486	0.38896	0.288394
StackLSTMmodel_weight_one_attribute_time-steps_14	0.218925	0.330796	0.328408	0.267264
StackLSTMmodel_weight_one_attribute_time-steps_21	0.220059	0.346851	0.387996	0.269111
GRU_model_weight_one_attribute_time-steps_7	0.221908	0.323945	0.363954	0.24998

- So sánh độ chính xác và MAE, MSE, RMSE theo từng Task của cuộc thi đề ra cho từng hạng mục dự báo.

Run ID	SubTask ID	Accuracy	Abs_difference
1	1	0.4	192.6
	2	0.4	8.5
	3	0.5	2
2	1	0.6	302.8
	2	0.4	8.5
	3	0.5	2

**Tài liệu  
tham  
khảo**

[1] Anh-Vu Mai-Nguyen, Van-Luon Tran, Minh-Son Dao, Koji Zettsu:

**Leverage the Predictive Power Score of Lifelog Data's Attributes to Predict the Expected Athlete Performance.** CLEF (Working Notes) 2020

[2] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Luca Piras, Michael Riegler, Pål Halvorsen, Mathias Lux, Minh-Triet Tran, Cathal Gurrin, Duc-Tien Dang-Nguyen:

**Overview of ImageCLEF Lifelog 2020: Lifelog Moment Retrieval and Sport Performance Lifelog.** CLEF (Working Notes) 2020

[3] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, Minh-Triet Tran, Tu-Khiem Le, Van-Tu Ninh, Cathal Gurrin:

**Overview of ImageCLEF lifelog 2019: Solve My Life Puzzle and Lifelog Moment Retrieval.** CLEF (Working Notes) 2019

[4] <https://www.aicrowd.com/challenges/imageclef-2020-lifelog-spl#data>

[5] <https://en.wikipedia.org/wiki/Lifelog>