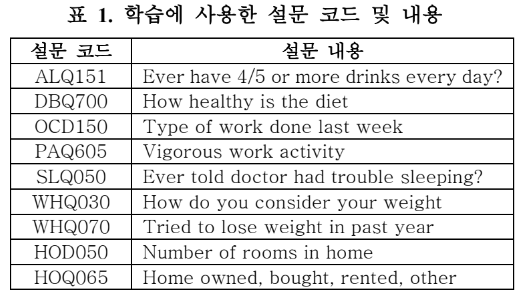
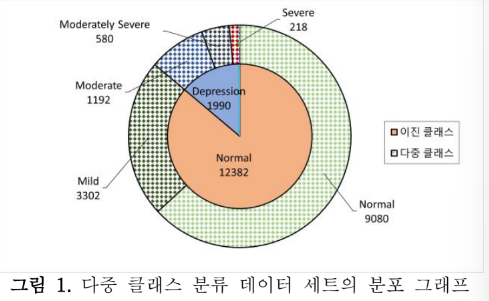
**NHANES 데이터를 이용한 생활 패턴 기반의 우울증 예측 머신러닝 기법 - 2021**



**데이터**

* 특성데이터: NHANES의 설문 데이터 중 표1에 있는 설문 내용.
* 분류 데이터: PHQ-9의 척도를 기준으로 각 우울증 정도를 사용.

→ 그림1과 같이 다중 클래스 / 이진 클래스로 분류



**학습 모델**: DNN, DT, KNN, MLP, GB, SVM, RF

**결과**

* 다중 클래스 분류: DNN이 55%의 F1 score로 가장 높음
* 이진 클래스 분류: DT가 71%의 F1 score로 가장 높음

**한계**: 불균형한 데이터셋

—

**기계학습 및 소셜미디어 이미지 기반 우울증 분석 모델 개발 - 2018**

**데이터**

Instagram의 *#depression*과 *#happy* 이미지 각 8001장, 9022장 크롤링

**모델 학습**

1. 색상정보 활용: 이미지에서 중요하다고 판단되는 부분에 가중치 → SVM, RF
2. 객체 정보 활용: Google Vision API로 객체 추출 → Counter Vector, TF-IDF
3. 딥러닝: CNN 기반의 Inception-v1을 이용해 학습

**결과**: 2번 방식이 92%의 정확도로 가장 높음.

**한계**: 학습 이미지의 수와 분류가 적었다는 점.

—

**우울증 치료를 위한 챗봇 개발의 방향성 고찰 - 2020**

**우울증 성향 연구 분석**

- TCI 성격 검사를 실시

→ 우울증 / 우울증 집단의 차이가 나타나는 6가지 척도(위험회피, 자극추구, 사회적 민감성, 자기 주도성, 연대감, 자기초월에 대한 성격 특성)을 찾아냈으며, 비관적, 잦은 피로감, 목적의식의 부재, 낮은 자신감 및 참을성 등과 같은 우울증 집단 성향 추정

- 역학연구 우울척도 (CESD)와 소셜미디어 사용률을 통해 초기 우울증 탐지 (Mok의 연구)

**우울증 탐색과 진단 AI**

- 불균형 소셜 미디어 데이터의 우울증 탐지: XA-BiLSTM 제안

- Fisher Vector 및 LTrP 알고리즘을 사용한 얼굴 이미지 우울증 감지 방법

- 통계 분석 및 자연어 처리를 통합한 AEKW 알고리즘

**우울증 진단 영역**

- 설문 기반의 1차적인 서비스를 통한 신속한 진단

- 사용자가 챗봇과 대화하면서 생산되는 데이터를 활용해 사전에 우울증을 탐지

- 사용자들의 SNS 비정형 데이터를 기반으로 AI 기술 활용

**챗봇의 치료 영역**

- 음악 / 웃음 / 인지행동치료(CBT) 등

**결론**

- 챗봇 개발 시 고려해야할 방법론에 대한 연구 부족

- 우울증을 치료하는 챗봇은 사실 기반 정보보다 일상 대화와 개인화된 서비스 지원이 필요

—

**부정적 감정 완화를 위한 BERGPT - 2021**

**선행연구**

*Replika*

* 대화할수록 챗봇이 업그레이드되며 사용자와의 대화를 기억.
* 재순위(Reranking) 모델을 사용해 응답하기 때문에 정확한 응답 가능.
* 선정된 후보 문장을 fine-tuning BERT를 통해 문장의 긍/부정 분류
* 계산된 긍정 비율 척도는 추후 학습 데이터로도 활용. 점수 산출에도 영향.

*BERT*

* 트랜스포머 구조에서 인코더(Encoder) 부분을 사용하는 모델.
* 양방향 모델과 WordPiece 토크나이징을 사용하는 것이 강점.

\* WordPiece 토크나이징: 문장에 들어있는 단어의 빈도수를 기바으로하여 단어 의미를 파악하는 방식

* 단어 간 관계 파악의 정확도가 높아 문맥 파악이 더 수월.

*KR-BERT*

* 기존의 음절 단위 토큰이 아닌 자소 단위의 토큰. (ex. 감 → ‘ㄱ’, ‘ㅏ’, ‘ㅁ’)
* 감정 분석, 질의 응답, 개체명 인식, 패러프레이즈 인식 과제에서 다른 한국어 BERT 모델보다 높거나 동등한 성능.

*KoGPT2*

* 음절 바이트 페어 인코딩과 트랜스포머의 디코더(decoder) 블록을 사용해 한국어 문장 학습.

**방법론**

제안 모델: KR-BERT와 KoGPT2-chatbot을 하나의 파이프라인으로 구성.

* 일상 대화 데이터를 사용하여 학습한 KR-BERT로 감정 분류 진행.

→ KoGPT2-chatbot의 추가 학습 데이터로 사용하여 모델 학습.

**데이터**

* Naver sentiment movie corpus v1.0
* 감성 대화 말뭉치 v1.0 (AI-Hub)
* 감성 분석 말뭉치 2020 v1.0 (국립국어원)
* 웰니스 대화 스크립트 (AI-Hub)
* 일상 대화 말뭉치 2020 v1.1 (국립국어원)

**결과**

제안한 모델(BERGPT-chatbot)이 기존의 KoGPT2-chatbot보다 뛰어난 성능

**한계점**

공개된 감정 데이터셋의 부족

—

**딥러닝을 이용한 음성신호 기반 우울증 진단 - 2021**

**사전연구**

* 스펙트로그램, 멜 스펙트로그램 특징을 적용해 CNN 모델을 통한 분류
* 원시 신호를 이용하여 LSTM 모델을 통해 분류

**데이터**

DAIC-WOZ: 불안, 우울증 및 PTSD 등과 같은 심리적 불안 상태 진단에 도움을 주기 위해 고안된 임상 인터뷰

+ pyAudioAnanlysis를 이용한 소음 제거

**모델**

*YAMNet*

* CNN 모델기반 사전 훈련 딥네트위크
* 음성신호를 시간-주파수 표현으로 변환하기 위해 Bark 스펙트로그램 특징이 사용.

*VGGish*

* 로그 멜 스펙트로그램 특징 추출 방법을 이용.

\* 로그 멜 스펙트로그램: 감정인식, 정신감정 상태 분류에 사용됨

**결과**

* YAMNet: 56.64% acc
* VGGish: 49.38% acc

**제안**

* 음성 데이터 추가 및 전처리를 통한 확장
* 표정이나 텍스트를 음성신호와 앙상블하는 기법 연구

—

**2D-CNN 기반 우울증 감지를 위한 음성데이터 전처리 - 2021**

**사전연구**

음성 세그먼트 특징 추출 시간과 인식률 사이의 균형에서 7초의 세그먼트 길이가 적합함을 확인

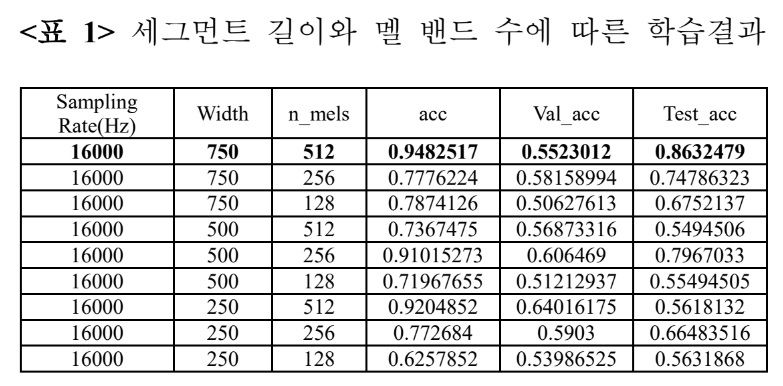
**데이터**: DAIC-WOZ

**전처리**: Librosa를 통한 Mel-Spectrogram 추출 (세그먼트 길이와 멜 밴드 수의 변화를 줌)

**모델**: 2D-CNN

**결과**

음성 세그먼트 길이 12sec(750 width), 멜 밴드 512일 때 가장 높은 정확도.



**제안**

ResNet50와 같은 이미지넷에서의 검증을 통해 최적의 세그먼트 길이 및 멜 밴드 수 확인

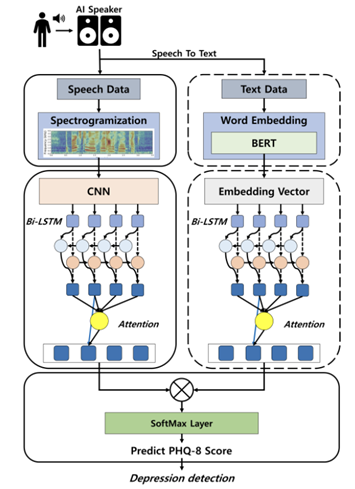
—

**AI 스피커를 활용한 어텐션 메커니즘 기반 멀티모달 우울증 감지 시스템(2021) – 박준희, 문남미**

<요약>

* 감성 분석을 통한 우울증 판별 (다중 데이터: 음성, 텍스트 → 멀티 모달)
* **음성, 텍스트 데이터** **수집**: 음성은 주파수 이미지로 변환하고 텍스트는 음성에서 따옴.
* cnn과 biLSTMM(Bidirectional Long Short-Term Memory Network) → 각 데이터의 특징 추출
* **텍스트 데이터**: BERT 알고리즘 기반 BiLSTM 사용, 음성 데이터: MFCC 기반 CNN-BiLSTM 모델 사용
* **어텐션 메커니즘**: self-attention을 적용하여 특징 벡터에 추가적인 가중치를 부여 → 높은 정확도의 우울증 판별
* **BERT**: 트랜스포머 인코더를 쌓아올린 구조. 문맥 정보를 가진 [CLS] 벡터를 따로 두어 층을 거칠 때마다 특징을 머금게 되는.., 서브워드 토크나이저(단어보다 더 작은 단위로) 사용하여 이미 알고 있는 단어가 나올 때까지 단어를 쪼갬.
* **LSTM** : RNN에서 학습하는 정보가 길어질수록 문제를 해결하기 어려워지는 문제점을 해결한 신경망. [[참고]](https://dgkim5360.tistory.com/entry/understanding-long-short-term-memory-lstm-kr)
* **BiLSTM** : 양방향 LSTM으로 두 개의 LSTM을 함께 사용하는 구조. 한 LSTM은 순차적으로 입력을 받는다면 나머지 하나는 거꾸로 입력을 받는다.
* 최종적으로 softmax 함수를 통해 우울증 판단
* 본 논문에서는 시스템을 설계, 제안만 하고 직접 우울증 점수를 도출한다든지 결과를 보이지 않음. 그러나 멀티모달 관련 논문을 많이 소개해주고 있음.

<시스템 개요 (보다 자세한 내용)>



1) AI 스피커로 데이터 수집

2) 임베딩(데이터를 벡터로 변환)

* **음성 데이터**: spectrogramization (시간에 따른 음압 → Fast Fourier Transform 푸리에 변환 → Mel-Spectrum → Mel-Frequency Cepstral coefficient 주파수 영역 이미지) → 2D CNN으로 특징 벡터 추출
* **텍스트 데이터**: 음성 데이터를 텍스트 데이터로 변환(Speech-To-Text) → 정규화 및 전처리 (NLTK 모듈, WordNetLemmatizer 사용) → BERT 알고리즘

3) 입력 데이터의 특징 학습

* **음성 데이터**: Bi-LSTM으로 학습, 학습 과정 중 어텐션(우울증 특징 정보) 수집. => CNN 특징 정보+ 어텐션
* **텍스트 데이터**: BiLSTM 모델로 학습. 음성 데이터와 마찬가지로 Self-Attention 사용

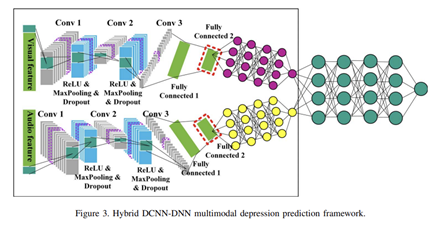
4) 우울증 점수 예측

* **두 데이터 통합** : Fully Connected Layer에서 통합. 이 임베딩 벡터를 softmax layer에 넣고 PHQ-8 스코어 예측. → **우울증 감지**

—

DCNN and DNN Based Multi-modal Depression Recognition(2017)

* 음성, 얼굴 영상, 텍스트 데이터 이용
* 위 논문의 참고 문헌 중 하나
* 남/여 데이터 따로 분석
* 음성, 얼굴 이미지에서 추출된 특징 → DCNN에 입력(high level 특징 학습) → DNN에서 phq-8 점수 예측. (단일 modal)
* 음성, 얼굴 이미지에서 추출된 특징 → DCNN-DNN에 입력(high level 특징 학습) → DNN에서 phq-8 점수 예측. + 얼굴 랜드마크의 범위와 속도를 정량화하는 표현 기법 “HDR” 제안 (멀티 modal)
* DCNN : 심층 컨볼루션 신경망. 컨볼루션+풀링(데이터의 공간 축약)이 번갈아가며.. 있는 구조. 마지막에 fully connected layer 있음.



* 2차원 얼굴 랜드마크에서 HDR(Histogram of Displacement Range)로 변환..수평/수직 방향에 대한 변위와 속도 계산?... HOG랑 비슷한 느낌..?
* 전체적으로 논문이 single-modal과 multi-modal의 성능을 따로따로 비교하는 형식.
* AVEC 2016 depression(<https://arxiv.org/pdf/1605.01600.pdf> ) dataset 이용. 이 데이터셋은 68개의 얼굴 랜드마크 좌표를 제공하고 있다. - 이게<https://dcapswoz.ict.usc.edu/> 이 데이터베이스를 썼다는 말일까..? (해당 데이터베이스에 대한 자세한 설명:<https://dcapswoz.ict.usc.edu/wp-content/uploads/2022/02/DAICWOZDepression_Documentation.pdf> )
* 각 데이터 부문 별로 phq-8 score와 우울증 여부 이진 분류값이 라벨링 되어있음.
* single-modal에서 선행연구 모델을 능가하는 결과가 나옴.
* multi-modal에서도 최신 모델보다 성능이 우수하다.

—

**Clinical Depression Detection in Adolescent by Face - 2018**

**목적**: 얼굴 이미지를 통한 우울증 / 비우울증 분류

**데이터셋**: AVEC 2013

**모델**: Fisher Vector 및 LTrP 알고리즘

**결과**: 최대 87.6% acc

—

**AVEC 2016- Depression, Mood, and Emotion Recognition Workshop and Challenge (2016)**

대회에서의 두 가지 과제

1) 우울증 분류 – DAIC-WOZ 데이터베이스(Ellie라는 가상의 인터뷰어와 인터뷰하는 비디오, 오디오 데이터 셋) 사용.

: 우울증 이진 분류 + 우울증 심각성 평가

2) 멀티모달 감정 인지 – RECOLA 데이터 베이스 사용

: 두 가지 정서 차원(arousal, valence)에 대한 회귀

1번 과제

* 우울증의 수준은 PHQ-8 점수로 정규화, 녹화별로 하나의 값으로 저장됨.
* PHQ-8 : 자기 평가 방법으로, 주관적이고 우울증 질문을 다룸. phq-9 질문과 비슷하지만 윤리적인 이유로 지워진 자살 생각에 대한 질문을 포함하고 있음.
* phq8 train&validation set(논문에서는 development set라고 했는데 같은걸까): 평균 우울 심각도 m=6.67

\* 비디오 데이터 어떻게 쓰는가

* OpenFace framework 사용하여 얼굴 랜드마크, 얼굴 영역 검출(HOG), 두 눈의 주시 방향, 얼굴의 방향(3D) 탐지
* AFECT(<https://www.mpt4u.com/> , 사람의 얼굴 표정을 수치로 나타내는 실시간 표정 인식 응용 프로그램)소프트웨어를 기반으로 감정(10가지)과 얼굴 행동 단위(AU, action unit, 얼굴 표정을 위한 낌새..? 섬세한 움직임인듯) 연속 측정
* 대회용으로는 제공한 것 같은데 raw 비디오는 사용불가하게 됐다고..

\*오디오 데이터 어떻게 쓰는가

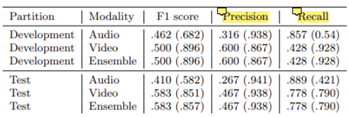
* COVAREP(말하기 분석을 위한 matlab, octave toolbox, object c&c++ 이용) : 화자의 음질, 운율 특성, 스펙트럼을 포착함. 우울증 탐지에 주로 쓰인다

결과

\*둘다 프레임 단위로 계산

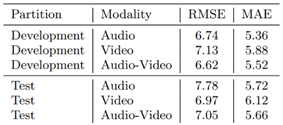
① 우울증 분류

* 스킷런 툴박스 이용하여 계산
* 확률적 경사하강법을 이용하여 linear svm에 fit하게 만들고 validation set으로 모델 검증
* 오디오&비디오 데이터는 validation set으로 grid search를 통해 최적의 하이퍼 파라미터 찾아냄.
* 오디오와 비디오의 융합(?)은 AND 논리로 이진 계산
* 테스트 성능, 괄호안은 not depressed class, 모두 평균값



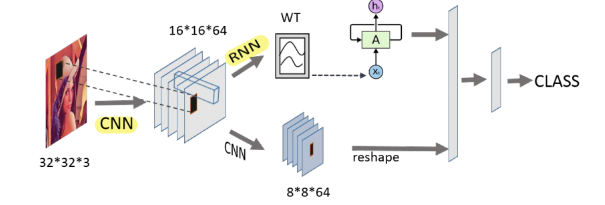
② 우울증 심각도 예측

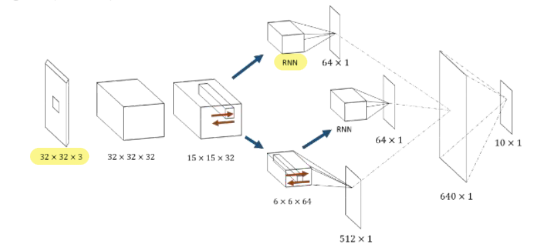
* 랜덤 포레스트를 사용하여 회귀 기준선 계산
* 하이퍼 파라미터는 트리의 숫자
* 테스트 성능: 예측값과 실제 phq-8 점수로 계산, 표 안의 값은 모든 프레임 평균값

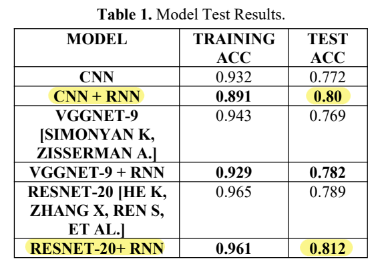


—

**CNN and RNN mixed model for image classification (2018)**







—

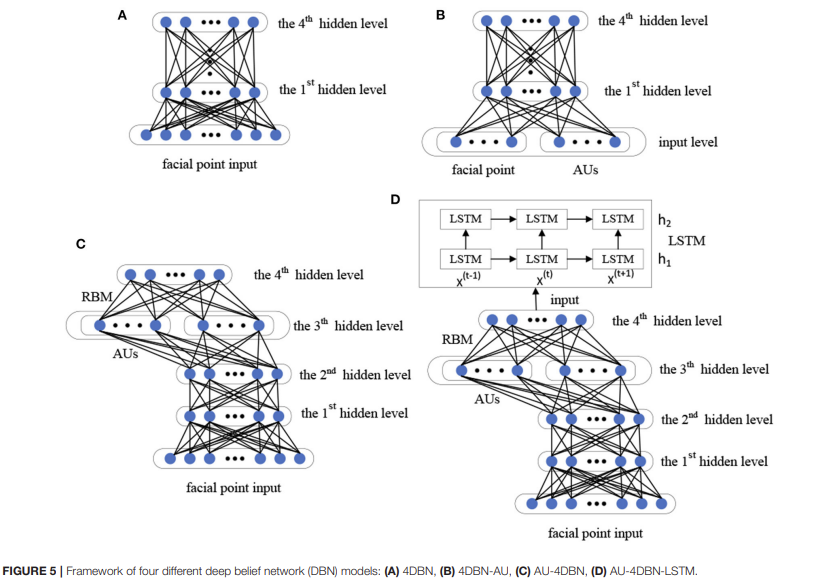
**Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks (2021)**

데이터셋: 중국인들 대상으로 직접 인터뷰하여 영상 데이터셋 제작. 랜드마크, AU 이외에도 pose, illumination, angle, skin color, resolution power 데이터도 추가

DBN(Deep Belief Network, 4가지 RBM 결합한 형태) 기반의 잠재적 우울증 위험 인식 모델 2가지 결합

(1) 2D 랜드마크+pose~resolution power 고려 모델(얼굴 외적인 요소만 고려), DBN이용

(2) 3D 랜드마크+AU : DBN+LSTM 이용



우울집단과 비우울 집단에게 5가지 자극을 주고, 그때의 valence와 arousal 차이 비교

결과: 3D-DGDN(모델 2)이 좀 더 높은 정확도 보임, 영화 시청, 표정 사진 보여주기 자극에서는 높은 정확도 보였지만, 질문을 던졌을 때 대답을 하는 얼굴 표정까지 녹화되어 정확도가 떨어졌다.

**Automatic Depression Detection via Learning and Fusing Features from Visual Cues (2022, arxiv)**

\* 우리 연구 중 이미지 처리 부분과 굉장히 흡사한 연구인 듯.. 똑같이 daic-woz 사용하고 우울증 여부 판단.

1. 서론

우리와 똑같은 고민을 하였다. 일시적인 행동만으로 우울증을 판단하기 어려울 것이라는 것! 그래서 temporal 정보를 고려할 필요가 있었다. 그러나 LSTM과 TCN 같은 건 문제점이 있었다.

첫 번째 문제점: 엄청나게 긴 시퀀스를 처리하기엔 살짝 적절하지 않음.

두 번째 문제점: 하나의 단서(레이블)를 처리하는데 유리, 여러가지 단서 처리할 때 서로의 상보성(서로 영향을 주고 받는 관계)을 무시하기 때문.

(용어 정리)

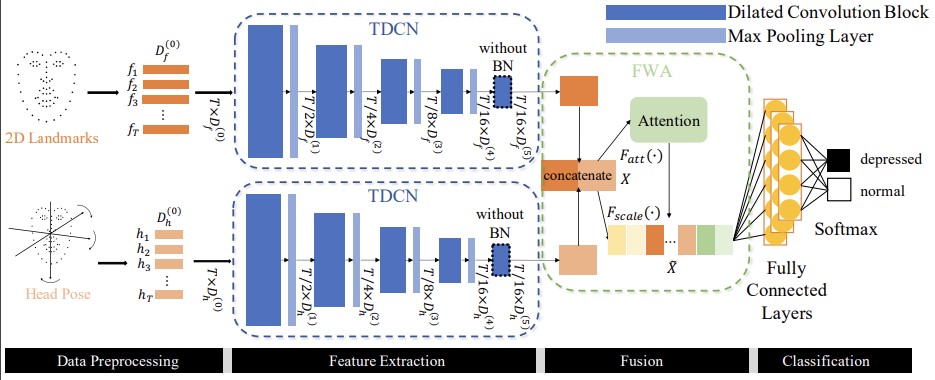
\* Temporal Dilated Convolutional Network (TDCN)

: Dilated Convolution Blocks가 여러 개 쌓여있는 모습. 이걸 사용하면 장기간 정보를 차례대로 학습하는데 유용한 듯.

\* Dilated Convolution에 대한 설명 : <https://zzsza.github.io/data/2018/02/23/introduction-convolution/> real-time segmentation에 자주 사용됨.

\* Feature-Wise Attention module (FWA) : 어텐션이 들어간 만큼... 중요한 특징이 보이면 가중치를 더 추가하는 역할인듯..

!제안하는 프레임워크!

TDCN+FWA

* TDCN의 역할: 우울증의 관련 정보를 추출한다.
* FWA의 역할: 특징 채널에 서로 다른 가중치를 할당하여 특징을 더 살린다.

TDCN은 두 가지 평행한 모듈이 우울증 탐지에 유용한 시간적 정보를 학습하게끔 하고, max pooling 층을 만들어서 overlength 문제를 해결하고자 하였다. 2D Landmark랑 Head Pose 모델을 따로 둔 듯.

FWA 모듈은 TDCN에서 학습한 특징들을 융합하고, 중요한 특징들을 강조하게 해서 전체적인 정확도를 높이게끔 한다. 마지막엔 그 특징들을 fully connected layers - softmax layer에 집어 넣어 분류(우울/비우울)를 시도한다.

\*본 연구에서 overlength 문제를 해결하기 위한 방법은?

1) 계산 복잡도를 줄이기 위해 고정된 크기를 가진 여러가지 시퀀스 샘플로 나눈다.. -> 우리 연구에선....시퀀스 크기가 다양할 것

2) 다양한 크기를 가진 지각 분야의 정보를 충분히 활용하는 TDCN 모델을 제안한다. (다양한 크기를 가진 곳이 어딘지는 모르겠다. 아마도 DCB의 특징차원 얘기하는 것 같기도...)

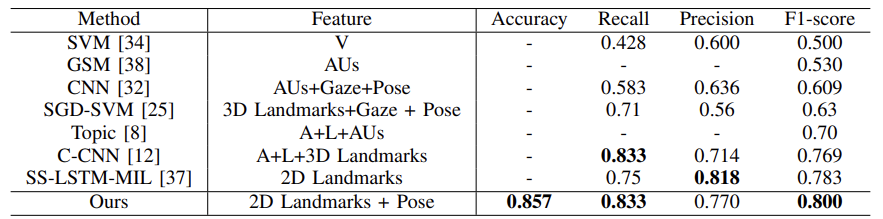
\*TDCN과 FWA의 구조와 원리

* TDCN : 5개의 Dilated Convolutional Blocks/ 4개의 max pooling layers 번갈아가며 있는 구조. 그래서 우울증 관련 정보를 다양한 규모에서 추출할 수 있는 것이다.
* DCB 역할: 다양한 지각 범위에서 유용한 정보들을 탐색한다.
* max pooling layer 역할: 특징 해상도를 계속 축소하고, 가장 중요한 결과를 조금씩 추출한다. + 시퀀스 길이 줄여서 모델의 복잡도 줄이기
* FWA : 원리가 어떻게 되는지는 잘 모르겠다... 먼저 TDCN에서 나온 두 결과를 합치고 global average pooling... 그리고 활성화함수에 넣는다..?

\* 데이터 전처리 : 본 논문에서도 똑같이 DAIC-WOZ 데이터셋 사용, input을 같은 크기로 두기 위해서 시퀀스 길이를 통일할 필요가 있었다. 그래서 head-first re-sample 실행해서 모든 샘플을 5000으로 길이를 맞췄다.

!결과!

1) Method+feature 측면



2) feature 측면.. : 2D 랜드마크와 pose를 쓴 게 가장 좋았다. 적당히 두 개만!

