<p dir="rtl">السيد الأستاذ الدكتور/ مقرر اللجنة العلمية الدائمة لترقية الأساتذة والأساتذة المساعدين للحاسبات والمعلومات</p>

<p dir="rtl">تحية طيبة وبعد – احيط سيادتكم علما بان البحث رقم ١ بياناتة كالتالي:</p>

<p dir="rtl"><u>عنوان البــــــحث باللغة الانجليزية:</u></p>

<h2 align="center">Bimodal Variational Autoencoder for Audiovisual Speech Recognition</h2>

<p dir="rtl"><u>مكان النشر و تاريخه:</u></p>

<p dir="rtl"><u>أسماء المشاركين في البحث:</u></p>

<p align="center">Hadeer M. Sayed, Hesham E. ElDeeb, Shereen A. Taie</p>

<p dir="rtl"><u>ملخص البــــــحث باللغة الانجليزية :</u></p>

Multimodal fusion is the idea of combining information in a joint representation of multiple modalities. The goal of multimodal fusion is to improve the accuracy of results from classification or regression tasks. This paper proposes a Bimodal Variational Autoencoder (BiVAE) model for audiovisual features fusion. Reliance on audiovisual signals in a speech recognition task increases the recognition accuracy, especially when an audio signal is corrupted. The BiVAE model is trained and validated on the CUAVE dataset. Three classifiers have evaluated the fused audiovisual features: Long-short Term Memory, Deep Neural Network, and Support Vector Machine. The experiment involves the evaluation of the fused features in the case of whether two modalities are available or there is only one modality available (i.e., cross-modality). The experimental results display the superiority of the proposed model (BiVAE) of audiovisual features fusion over the state-of-the-art models by an average accuracy difference $\simeq$ 3.28% and 13.28% for clean and noisy, respectively. Additionally, BiVAE outperforms the state-of-the-art models in the case of cross-modality by an accuracy difference $\simeq$ 2.79% when the only audio signal is available and 1.88% when the only video signal is available. Furthermore, SVM satisfies the best recognition accuracy compared with other classifiers.