TABLE V: Performance on EDAQA dataset, Study 1

| | | Study1 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | s0 | s1 | s2 | s3 | s4 | s5 | Average |
| Rad Original | Acc: | 0.863 | 0.957 | 0.965 | 0.865 | 0.960 | 0.732 | 0.890 ± 0.09 |
| | AUC: | 0.869 | 0.925 | 0.870 | 0.922 | 0.914 | 0.878 | 0.896 ± 0.06 |
| | F1: | 0.577 | 0.118 | 0.461 | 0.777 | 0.665 | 0.695 | 0.549 ± 0.24 |
| Rad Bayes | Acc: | 0.858 | 0.962 | 0.967 | 0.810 | 0.975 | 0.660 | 0.872 ± 0.12 |
| | AUC: | 0.841 | 0.873 | 0.883 | 0.848 | 0.985 | 0.914 | 0.891 ± 0.11 |
| | F1: | 0.502 | 0.115 | 0.436 | 0.606 | 0.788 | 0.566 | 0.502 ± 0.22 |
| WiderNet 2x | Acc: | 0.867 | 0.973 | 0.972 | 0.880 | 0.950 | 0.779 | 0.904 ± 0.08 |
| | AUC: | 0.833 | 0.908 | 0.896 | 0.946 | 0.859 | 0.929 | 0.895 ± 0.04 |
| | F1: | 0.585 | 0.192 | 0.570 | 0.803 | 0.577 | 0.777 | 0.584 ± 0.22 |
| WiderNet 2x Bayes | Acc: | 0.881 | 0.969 | 0.969 | 0.871 | 0.977 | 0.805 | 0.912 ± 0.07 |
| | AUC: | 0.884 | 0.961 | 0.945 | 0.923 | 0.987 | 0.943 | 0.941 ± 0.03 |
| | F1: | 0.620 | 0.156 | 0.489 | 0.798 | 0.806 | 0.805 | 0.612 ± 0.26 |
| WiderNet 4x | Acc: | 0.865 | 0.978 | 0.975 | 0.882 | 0.948 | 0.783 | 0.905 ± 0.08 |
| | AUC: | 0.827 | 0.882 | 0.898 | 0.945 | 0.852 | 0.907 | 0.885 ± 0.06 |
| | F1: | 0.570 | 0.183 | 0.615 | 0.805 | 0.566 | 0.781 | 0.587 ± 0.22 |
| WiderNet 4x Bayes | Acc: | 0.885 | 0.979 | 0.973 | 0.877 | 0.975 | 0.813 | 0.917 ± 0.07 |
| | AUC: | 0.894 | 0.977 | 0.958 | 0.933 | 0.975 | 0.950 | **0.948 ± 0.11** |
| | F1: | 0.635 | 0.219 | 0.610 | 0.811 | 0.783 | 0.815 | 0.645 ± 0.23 |
| WiderNet 8x | Acc: | 0.866 | 0.982 | 0.876 | 0.881 | 0.900 | 0.757 | 0.877 ± 0.07 |
| | AUC: | 0.821 | 0.853 | 0.874 | 0.949 | 0.815 | 0.903 | 0.869 ± 0.05 |
| | F1: | 0.585 | 0.176 | 0.520 | 0.802 | 0.503 | 0.736 | 0.554 ± 0.22 |
| WiderNet 8x Bayes | Acc: | 0.888 | 0.984 | 0.974 | 0.875 | 0.973 | 0.824 | **0.920 ± 0.07** |
| | AUC: | 0.894 | 0.980 | 0.951 | 0.933 | 0.961 | 0.954 | 0.945 ± 0.03 |
| | F1: | 0.645 | 0.269 | 0.612 | 0.809 | 0.768 | 0.828 | **0.655 ± 0.21** |
| WiderNet 16x | Acc: | 0.634 | 0.933 | 0.940 | 0.826 | 0.921 | 0.727 | 0.830 ± 0.13 |
| | AUC: | 0.775 | 0.805 | 0.877 | 0.922 | 0.798 | 0.854 | 0.838 ± 0.06 |
| | F1: | 0.482 | 0.175 | 0.552 | 0.748 | 0.500 | 0.719 | 0.529 ± 0.21 |
| WiderNet 16x Bayes | Acc: | 0.883 | 0.983 | 0.973 | 0.880 | 0.974 | 0.823 | 0.919 ± 0.07 |
| | AUC: | 0.871 | 0.986 | 0.953 | 0.936 | 0.961 | 0.950 | 0.943 ± 0.04 |
| | F1: | 0.627 | 0.272 | 0.601 | 0.819 | 0.775 | 0.827 | 0.654 ± 0.21 |

TABLE VI: Performance on EDAQA dataset, Study 2

| | | Study2 | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | s0 | s1 | s2 | s3 | s4 | s5 | Average |
| Rad Original | Acc: | 0.863 | 0.962 | 0.967 | 0.921 | 0.958 | 0.767 | 0.906 ± 0.08 |
| | AUC: | 0.877 | 0.907 | 0.881 | 0.965 | 0.911 | 0.894 | 0.906 ± 0.08 |
| | F1: | 0.563 | 0.130 | 0.533 | 0.884 | 0.643 | 0.727 | 0.580 ± 0.25 |
| Rad Bayes | Acc: | 0.868 | 0.960 | 0.967 | 0.861 | 0.981 | 0.652 | 0.881 ± 0.12 |
| | AUC: | 0.866 | 0.931 | 0.945 | 0.904 | 0.991 | 0.884 | 0.920 ± 0.05 |
| | F1: | 0.538 | 0.098 | 0.463 | 0.740 | 0.839 | 0.502 | 0.530 ± 0.26 |
| WiderNet 2x | Acc: | 0.874 | 0.982 | 0.974 | 0.913 | 0.960 | 0.833 | 0.923 ± 0.06 |
| | AUC: | 0.866 | 0.938 | 0.924 | 0.961 | 0.899 | 0.941 | 0.921 ± 0.03 |
| | F1: | 0.610 | 0.234 | 0.615 | 0.869 | 0.646 | 0.840 | 0.636 ± 0.23 |
| WiderNet 2x Bayes | Acc: | 0.890 | 0.974 | 0.979 | 0.899 | 0.981 | 0.858 | 0.930 ± 0.05 |
| | AUC: | 0.905 | 0.972 | 0.969 | 0.949 | 0.992 | 0.958 | 0.958 ± 0.03 |
| | F1: | 0.660 | 0.203 | 0.660 | 0.854 | 0.833 | 0.866 | 0.679 ± 0.25 |
| WiderNet 4x | Acc: | 0.883 | 0.986 | 0.978 | 0.919 | 0.955 | 0.808 | 0.921 ± 0.07 |
| | AUC: | 0.864 | 0.961 | 0.919 | 0.966 | 0.880 | 0.933 | 0.920 ± 0.08 |
| | F1: | 0.662 | 0.288 | 0.664 | 0.876 | 0.618 | 0.812 | 0.653 ± 0.20 |
| WiderNet 4x Bayes | Acc: | 0.894 | 0.982 | 0.978 | 0.900 | 0.978 | 0.864 | **0.933 ± 0.05** |
| | AUC: | 0.913 | 0.982 | 0.967 | 0.954 | 0.988 | 0.962 | **0.961 ± 0.03** |
| | F1: | 0.672 | 0.277 | 0.680 | 0.856 | 0.813 | 0.873 | 0.695 ± 0.22 |
| WiderNet 8x | Acc: | 0.871 | 0.973 | 0.970 | 0.910 | 0.955 | 0.739 | 0.903 ± 0.09 |
| | AUC: | 0.839 | 0.904 | 0.908 | 0.962 | 0.885 | 0.920 | 0.903 ± 0.03 |
| | F1: | 0.586 | 0.222 | 0.641 | 0.861 | 0.606 | 0.700 | 0.602 ± 0.21 |
| WiderNet 8x Bayes | Acc: | 0.889 | 0.984 | 0.978 | 0.900 | 0.979 | 0.871 | **0.933 ± 0.05** |
| | AUC: | 0.908 | 0.991 | 0.965 | 0.951 | 0.986 | 0.960 | 0.960 ± 0.03 |
| | F1: | 0.649 | 0.329 | 0.682 | 0.855 | 0.821 | 0.880 | **0.703 ± 0.21** |
| WiderNet 16x | Acc: | 0.858 | 0.954 | 0.976 | 0.850 | 0.927 | 0.721 | 0.881 ± 0.09 |
| | AUC: | 0.811 | 0.930 | 0.920 | 0.928 | 0.869 | 0.893 | 0.892 ± 0.05 |
| | F1: | 0.499 | 0.256 | 0.665 | 0.774 | 0.547 | 0.681 | 0.570 ± 0.18 |
| WiderNet 16x Bayes | Acc: | 0.890 | 0.987 | 0.978 | 0.899 | 0.978 | 0.860 | 0.932 ± 0.05 |
| | AUC: | 0.899 | 0.990 | 0.951 | 0.949 | 0.975 | 0.961 | 0.954 ± 0.03 |
| | F1: | 0.657 | 0.356 | 0.660 | 0.853 | 0.808 | 0.869 | 0.700 ± 0.19 |

TABLE VII: Performance on ESDB dataset

| | | ESDB | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | s0 | s1 | s2 | s3 | s4 | s5 | s6 | Average |
| Rad Original | Acc: | 0.837 | 0.84 | 0.863 | 0.837 | 0.889 | 0.842 | 0.806 | 0.845 ± 0.03 |
| | AUC: | 0.922 | 0.913 | 0.926 | 0.917 | 0.954 | 0.906 | 0.878 | 0.916 ± 0.02 |
| | F1: | 0.83 | 0.671 | 0.9 | 0.866 | 0.918 | 0.862 | 0.844 | 0.841 ± 0.08 |
| Rad Bayes | Acc: | 0.78 | 0.892 | 0.864 | 0.862 | 0.909 | 0.885 | 0.817 | 0.858 ± 0.05 |
| | AUC: | 0.881 | 0.94 | 0.917 | 0.941 | 0.962 | 0.945 | 0.888 | 0.925 ± 0.03 |
| | F1: | 0.75 | 0.741 | 0.903 | 0.883 | 0.935 | 0.896 | 0.852 | 0.852 ± 0.08 |
| WiderNet 2x | Acc: | 0.855 | 0.867 | 0.867 | 0.847 | 0.901 | 0.87 | 0.826 | 0.862 ± 0.02 |
| | AUC: | 0.934 | 0.919 | 0.937 | 0.928 | 0.964 | 0.931 | 0.895 | 0.930 ± 0.02 |
| | F1: | 0.844 | 0.701 | 0.902 | 0.872 | 0.927 | 0.887 | 0.859 | 0.856 ± 0.07 |
| WiderNet 2x Bayes | Acc: | 0.803 | 0.911 | 0.863 | 0.878 | 0.927 | 0.914 | 0.856 | 0.879 ± 0.04 |
| | AUC: | 0.913 | 0.95 | 0.928 | 0.954 | 0.975 | 0.965 | 0.926 | 0.944 ± 0.02 |
| | F1: | 0.767 | 0.768 | 0.901 | 0.895 | 0.947 | 0.924 | 0.883 | 0.869 ± 0.07 |
| WiderNet 4x | Acc: | 0.857 | 0.868 | 0.866 | 0.85 | 0.904 | 0.87 | 0.825 | 0.863 ± 0.02 |
| | AUC: | 0.935 | 0.92 | 0.937 | 0.93 | 0.967 | 0.931 | 0.893 | 0.930 ± 0.02 |
| | F1: | 0.846 | 0.7 | 0.901 | 0.874 | 0.93 | 0.887 | 0.858 | 0.857 ± 0.07 |
| WiderNet 4x Bayes | Acc: | 0.812 | 0.912 | 0.86 | 0.88 | 0.927 | 0.916 | 0.851 | 0.880 ± 0.04 |
| | AUC: | 0.921 | 0.949 | 0.928 | 0.955 | 0.976 | 0.967 | 0.924 | 0.946 ± 0.02 |
| | F1: | 0.774 | 0.767 | 0.899 | 0.898 | 0.948 | 0.925 | 0.879 | 0.870 ± 0.07 |
| WiderNet 8x | Acc: | 0.859 | 0.87 | 0.862 | 0.854 | 0.907 | 0.872 | 0.814 | 0.863 ± 0.03 |
| | AUC: | 0.933 | 0.918 | 0.936 | 0.931 | 0.968 | 0.933 | 0.887 | 0.929 ± 0.02 |
| | F1: | 0.848 | 0.696 | 0.897 | 0.878 | 0.932 | 0.888 | 0.849 | 0.856 ± 0.08 |
| WiderNet 8x Bayes | Acc: | 0.815 | 0.913 | 0.858 | 0.885 | 0.929 | 0.916 | 0.846 | **0.881 ± 0.04** |
| | AUC: | 0.925 | 0.948 | 0.928 | 0.958 | 0.978 | 0.968 | 0.923 | **0.947 ± 0.02** |
| | F1: | 0.778 | 0.764 | 0.896 | 0.901 | 0.949 | 0.925 | 0.874 | **0.870 ± 0.07** |
| WiderNet 16x | Acc: | 0.858 | 0.87 | 0.857 | 0.856 | 0.912 | 0.859 | 0.806 | 0.860 ± 0.03 |
| | AUC: | 0.931 | 0.913 | 0.933 | 0.93 | 0.969 | 0.924 | 0.881 | 0.926 ± 0.03 |
| | F1: | 0.847 | 0.696 | 0.893 | 0.879 | 0.935 | 0.874 | 0.841 | 0.852 ± 0.08 |
| WiderNet 16x Bayes | Acc: | 0.816 | 0.912 | 0.853 | 0.888 | 0.93 | 0.907 | 0.838 | 0.878 ± 0.04 |
| | AUC: | 0.929 | 0.945 | 0.923 | 0.959 | 0.978 | 0.965 | 0.919 | 0.945 ± 0.02 |
| | F1: | 0.776 | 0.758 | 0.892 | 0.904 | 0.95 | 0.916 | 0.866 | 0.866 ± 0.07 |