



C 509
Mars 2020

Klassificering av ostronyngel med hjälp av artificiell intelligens

Jens Wilhelmsson, Åsa Strand, Anna-Lisa Wrangé, Karl Hunter, Torbjörn Johansson



I samarbete med: Sara Meurling, Uppsala Universitet och Pierre de Wit, Göteborgs universitet

Författare: Jens Wilhelmsson, Åsa Strand, Anna-Lisa Wrangle, Karl Hunter, Torbjörn Johansson, IVL Svenska Miljöinstitutet

Medel från: Livsmedelsstrategin via Jordbruksverket och MarGen II (Interreg Dnr 6.6.6 – Ä 2019-557, samt från EUs H2020 forsknings och innovationsprogram (Grant agreement No 818173.)

Fotograf: Åsa Strand

Rapportnummer C509, Mars 2020

ISBN 978-91-7883-167-8

Upplaga Finns endast som PDF-fil för egen utskrift

© **IVL Svenska Miljöinstitutet 2020**

IVL Svenska Miljöinstitutet AB, Box 210 60, 100 31 Stockholm

Tel 010-788 65 00 // www.ivl.se

Rapporten har granskats och godkänts i enlighet med IVL:s ledningssystem

Innehållsförteckning

Summary	3
Sammanfattning.....	4
1 Inledning	5
1.1 Bakgrund	5
2 Metod	8
2.1 Datainsamling.....	8
2.2 Genetiska analyser	9
2.3 Bildanalys.....	11
2.3.1 Identifiering av ostron	11
2.3.2 Klassificering av ostron	13
3 Resultat.....	15
3.1 Fotografering 1	15
3.1.1 Modell tränad på dataset 1.....	15
3.1.2 Modell tränad på dataset 2.....	18
3.2 Fotografering 2	19
3.3 Artbestämning genom DNA-analys	22
4 Diskussion och slutsats	23
5 Referenser.....	25

Summary

Mussels and oysters are healthy and sustainably produced foods with great economic value. One of the most valuable bivalve molluscs in Sweden is the flat oyster, *Ostrea edulis*. However, domestic production of oysters fails to meet the demand on the local market at the same time as the Swedish aquaculture industry has difficulties expanding due to limited availability of oyster spat.

Traditionally in Sweden, spat for aquaculture have been collected with the help of sea-based collectors, but since the introduction and establishment of the Pacific oyster (*Magallana gigas*) in 2006, the possibility of collecting spat with this field-based technology has drastically decreased. Both flat oyster and Pacific oysters attach to the collectors. Since aquaculture of Pacific oysters is not allowed in Sweden due to that the species is classified as an invasive species, the collected oyster spat must be sorted by species and all Pacific oysters must be destroyed, which is neither practical nor economically feasible for the industry today. The overall aim of the project is therefore to contribute to an increase in Swedish oyster cultivation by facilitating the growers to access flat oyster spat.

To achieve this goal, an algorithm has been developed that can classify images of oysters as either Pacific oysters or flat oysters through image processing and machine learning. The algorithm is based on supervised machine learning, which is a category in machine learning where the algorithm is given images and associated facts with what the image represents. In this project, the pictures were of oysters and the fact their species identification. Over 1,000 images of oysters were used to improve the algorithm's performance. The images were divided into 80% training data and 20% validation data. Training data was used to train the algorithm and the performance of the algorithm was evaluated based on validation data not used during the training procedure. In addition to oysters that were easily identified by species, as Pacific or flat oysters, 200 oysters were included that had unclear species identity based on initial visual assessment. These oysters were genetically analysed to determine species identity. The result shows that the vast majority of oysters were classified correctly in all image analyses (98.4%). Even the model trained on only clearly identified oysters, classified 95.9% of the oysters correctly in a validation data that included oysters that could not be determined easily by eye, but needed genetic verification. Of those oysters that were incorrectly classified in the image analysis, all were flat oysters classified as Pacific oysters. Of the initially "uncertain" oysters, it was verified (through genetic analysis) that all belonged to the native flat species *Ostrea edulis*. The result shows good potential for the development of an oyster sorting machine, which would enable an increase in the supply of flat oyster spat based on collection with sea-based collectors in Sweden.

Sammanfattning

Musslor och ostron är hälsosamma och hållbart producerade livsmedel med stort ekonomiskt värde. Ett av de mest värdefulla tvåskaliga blötdjuren i Sverige är det platta ostronet, *Ostrea edulis*. Den inhemska produktionen av ostron täcker dock inte efterfrågan på den lokala marknaden samtidigt som den svenska vattenbruksindustrin har svårt att expandera på grund av brist på ostronyngel. Traditionellt sett i Sverige har yngel för odling samlats in med hjälp av havsbaserade kollektorer, men sedan etableringen av det främmande arten stillahavsostron (*Magallana gigas*) 2006 så har möjligheten till insamling av yngel med denna fältnbaserade teknik försämrats drastiskt. På kollektorerna fäster sig numera inte bara det platta ostronet utan också yngel av stillahavsostron. Då odling av stillahavsostron inte är tillåtet i Sverige pga. att arten klassas som en invasiv art, måste de insamlade ostronynglena sorteras med avseende på art och alla stillahavsostronyngel måste destrueras, vilket varken är praktiskt eller ekonomiskt möjligt. Projektets övergripande syfte är därför att bidra till en ökning av den svenska ostronodlingen genom att underlätta för odlarna att få tag på yngel av det inhemska platta ostronet.

För att uppnå detta mål har en algoritm utvecklats som genom bildbehandling och maskininlärning kan klassificera bilder av ostron som antingen Stillahavsostron eller platta ostron. Algoritmen bygger på övervakad maskininlärning, vilket är en kategori inom maskininlärning där algoritmen ges bilder samt tillhörande facit med vad bilden representerar. I det här projektet utgjordes bilderna av ostron och facit av arttillhörighet. För att algoritmens prestanda skulle bli bra användes över 1 000 bilder av ostron. Bilderna delades upp i 80 % träningsdata och 20 % valideringsdata. Träningsdata användes för att träna upp algoritmen och algoritmens prestanda utvärderades utifrån valideringsdata som inte använts under träningsförfarandet. Förutom tydliga ostron av båda arter inkluderades 200 ostron som hade otydlig arttillhörighet baserat på initial visuell bedömning. Dessa ostron analyserades genetiskt för att fastställa arttillhörighet. Resultatet visar att de allra flesta ostronen klassificerades till korrekt art i samtliga bildanalyser (98,4 %). Till och med modellen som tränades på endast säkra ostron klassificerade 95,9 % av ostronen korrekt i ett valideringsdata som inkluderade ostron som inte kunde artbestämmas okulärt utan verifierades genom DNA-analys. Av de som klassificerats felaktigt i bildanalysen utgjordes samtliga av *Ostrea edulis* som klassificerats som *Magallana gigas*. Av de initialt "osäkra" ostronen verifierades (genom genetisk analys) att alla tillhörde den inhemska arten *Ostrea edulis*. Resultatet visar på god potential för utvecklingen av en ostronsorterare vilken skulle möjliggöra en ökning av tillgången på ostronyngel baserat på insamling med havsbaserade kollektorer i svenska havsområden.

1 Inledning

Detta projekt har finansierats av Jordbruksverket genom riktade satsningar inom Livsmedelsstrategin (2016). Det övergripande målet för livsmedelsstrategin är att Sverige ska ha en konkurrenskraftig livsmedelskedja där den totala livsmedelsproduktionen ökar samtidigt som relevanta nationella miljömål nås, med syftet att skapa tillväxt och sysselsättning och bidra till hållbar utveckling i hela landet. Inom ett av de tre strategiska målen för att uppnå denna vision lyfts behovet av stöd till kunskaps- och innovationssystemet för att bidra till ökad produktivitet och innovation i livsmedelskedjan samt till hållbar produktion och konsumtion av livsmedel.

Vattenbruk, speciellt odling av arter på låga trofiska nivåer som t.ex. alger, musslor och ostron lyfts internationellt ofta som en framtidsnäring (Fish to 2030, 2013; Hilborn et al. 2018) som kan bidra till en hållbar expansion av livsmedelsproduktionen globalt. Detta noteras också inom Vattenbruksstrategin (2012) där odling av skaldjur (musslor och ostron) nämns som en näring med stor potential för expansion av vattenbruk på den svenska västkusten.

Trots en hög efterfrågan och stor expansionspotential står den svenska ostronnäringen inför ett antal hinder som begränsar näringens utveckling. Ett av dessa hinder utgörs av brist på yngel av det inhemska ostronet *Ostrea edulis*. Detta projekts övergripande syfte är därför att bidra till en ökning av den svenska ostronnäringen genom att öka tillgången till yngel av det inhemska ostronet. Detta görs genom utveckling av ett bildidentifieringsprogram som kan särskilja yngel av de inhemska platta ostronen (*Ostrea edulis*) från yngel av det invasiva stillahavsostronet (*Magallana gigas*) vid yngelsamling baserad på havsbaserade kollektorer.

1.1 Bakgrund

Den beräknade befolkningmängden på 9,7 miljarder människor år 2050 (UN 2019) kommer att öka belastningen på existerande naturresurser. En stor utmaning är att tillgodose dessa människor med mat (UN 2015), och vattenbruk förs ofta fram som en möjlig lösning för att möta denna utmaning. Arter på låga trofiska nivåer, t.ex. musslor och ostron (bivalver), är utmärkta arter för detta. De är goda att äta och är nyttiga, de är lätta att odla, och de har positiva miljöeffekter som att minska övergödning genom att de filtrerar ut partiklar och mikroalger som föda ur vattnet (Kotta et al. 2020).

En av de mest värdefulla marina vattenbruksarterna är det platta ostronet, *Ostrea edulis*. Idag skördas i Europa drygt 1 100 ton platta ostron från vilda populationer och odlas 2 000 ton årligen (FAO 2020). I jämförelse är produktionen av arten i Sverige blygsamma 9 ton, varav merparten kommer från skörd av vilda ostronbankar i havet (FAO 2020). Den inhemska produktionen av ostron täcker inte efterfrågan på den lokala marknaden, och årligen importeras ca 600 ton ostron (både *Ostrea edulis* och Stillahavsostreon [*Magallana gigas*, tidigare *Crassostrea gigas*]) till Sverige (SCB 2020).

I Europa har parasiter (*Bonamia ostreae* och *Martelia refringens*) och överfiske minskat produktionen av det platta ostronet kraftigt (Ospar 2008). Då dessa parasiter hittills inte påträffats i Sverige så finns goda förutsättningar för odling av *Ostrea edulis*, vilket återspeglas i ett starkt intresse från både befintliga och nya vattenbruksaktörer att utöka produktionen av arten. Trots det stora intresset för arten så är den svenska ostronindustrin liten och tillväxten är minimal. En av de

viktigaste faktorerna som begränsar expansionen av ostronnäringen i Sverige idag är tillgången till ostronyngel.

På grund av småskaligheten inom ostronnäringen i Sverige har ostronyngel för odling traditionellt sett samlats in med hjälp av havsbaserade kollektorer, små kalkade plastskivor som hängs ut i havet under sommaren när ostronen förökar sig (Figur 1). Sedan etablering av det främmande Stillahavsostronet (*M. gigas*) 2006 så har möjligheten till insamling av yngel av vår inhemska ostronart med denna teknik försämrats drastiskt. På kollektorerna bottenfäller (sätter sig) numera inte bara yngel av det platta ostronet utan också yngel av Stillahavsostron, som ofta dominerar till antalet. Då odling av Stillahavsostron inte är tillåtet eftersom det klassas som en invasiv art, måste ostronynglen sorteras med avseende på art och alla Stillahavsostronyngel måste destrueras. Detta är varken praktiskt eller ekonomiskt möjligt på grund av det stora antalet yngel som måste sorteras för att få fram en tillräcklig mängd yngel av platta ostron för odling, samt att ynglen både är väldigt små och i vissa fall svåra att särskilja till art. Endast i något enstaka område längs svenska västkusten kan tillräckligt mycket yngel av platta ostron samlas in med denna teknik för att underhålla en småskalig ostronproduktion, vilket har stoppat utvecklingen av odling av det inhemska platta ostronet i Sverige.



Figur 1. Havsbaserade yngelkollektorer för insamling av ostronyngel. Bild: Åsa Strand

För att stärka utvecklingen av det svenska vattenbruket, och framförallt ostronsektorn, krävs därför utveckling av metoder för att underlätta produktion av yngel av det inhemska platta ostronet. Detta projekt har därför som mål att ta fram ett program (en klassificeringsalgoritm) som baserat på bilder av ostronyngel kan avgöra vilken av arterna, *O. edulis* och *M. gigas*, som



ostronynglen på bilderna utgörs av. Programmet ska fungera så att det i framtiden ska kunna användas i en ostronsorterare som delar upp ostron baserat på vilken art de tillhör.

En sådan klassificeringsalgoritm kan bygga på en typ av metod som kallas övervakad maskininlärning. Skillnaden mellan övervakad och oövervakad maskininlärning är att man i övervakad maskininlärning medvetet säger till en algoritm vad man vill att den ska lära sig, medan man med oövervakad maskininlärning låter algoritmen gruppera data utan att specificera vilken information i data man vill att den ska gruppera efter. Eftersom uppgiften för algoritmen är känd, det vill säga gruppera ostron efter vilken art de tillhör, så kommer övervakad maskininlärning att användas i detta projekt. Neurala nätverk hör till kategorin övervakade maskininlärningsmetoder och är vanliga att använda vid uppgifter som har med datorseende att göra. Det är även den maskininlärningsmetod som uppnått bäst resultat i klassificeringstävlingar där algoritmer ställts mot varandra¹.

Det finns många olika arkitekturer av neurala nätverk och nya arkitekturer tas fram hela tiden. Som utgångspunkt i detta projekt har vi valt att utgå från en nätverksarkitektur som historiskt presterat bra på klassificeringsuppgifter. Att utgå från befintliga arkitekturer har fördelen att det går att hitta förtränade nätverk som man kan bygga vidare på, eftersom det är tids- och beräkningskraftskrävande att träna ett neuralt nätverk från grunden.

¹ http://host.robots.ox.ac.uk:8080/leaderboard/main_bootstrap.php

2 Metod

Projektet har bestått av två delar, datainsamling och bildanalys.

2.1 Datainsamling

Ostronyngel samlades in med hjälp av havsbaserade kollektorer under sommaren-hösten 2018 inom ett forskningsprojekt som finansieras av Europeiska havs och fiskerifonden (EHFF, förmedlat genom Jordbruksverket, Dnr. 2017-925). Kollektorerna placerades ut på 2–3 m djup och 4–6 m djup på fem olika platser mellan Koster och Lysekil i juli 2018. Kollektorerna skördades i oktober och november 2018, och ynglen överfördes till laboratoriet vid Kristinebergscentrat för rengöring och sortering. Ynglen sorterades initialt i fyra kategorier: "tydlig OE" (yngel som definitivt är *O. edulis*, OE), "osäkra OE" (yngel som liknar OE men avviker något från den allmänna bilden av hur OE ska se ut), "osäkra MG" (yngel som liknar *M. gigas*, MG, men avviker något från den allmänna bilden av hur MG ska se ut) och "tydlig MG" (yngel som definitivt är MG). Efter sortering placerades ynglen i odlingskorgar på en odlingsanläggning för forskningsbruk i Gullmarsfjorden strax utanför Kristineberg. De säkert artidentifierade OE- och MG-ynglen fotograferades vid den första sorteringen 2018. Ostronen som användes vid fotografering 1 var mellan 4 och 5 cm, med en våtvikt på 7-14 gram. Detta bildmaterial har gjorts tillgängligt till detta projekt och har använts som ett underlag för träning av ostronklassificeringsprogramvaran.

I oktober 2019 gjordes en fotodokumentation där 200 säkra OE (skalllängd 39-51 mm, våtvikt: 5-9 g), 200 säkra MG (41-53 mm, 9-19 g) och 200 individer med osäker arttillhörighet ("unsure", US, 37-52 mm, 5-11 g) fotograferades. Fotodokumentation utfördes genom att ynglen placerades på ett rutnät på vit bakgrund, utan överlapp mellan individer. Bilder togs av båda sidorna av ynglen. Fotografierna togs med en kamera monterad på ett stativ (Canon EOS 200D med ett Canon EF-S 60mm f/2.8 Macro objektiv). Varje individ fick ett ID-nummer och biometridata för varje individ (längd/bredd/djup/våtvikt) samlades in som ett komplement till bilderna.

I november 2019 gjordes en ytterligare, kompletterande fotografering. Den här gången fotograferades ett större antal ostron i samma bild, utan någon slags rutnät eller indelning. Det enda kravet den här gången var att bakgrunden skulle vara vit samt att inga ostron skulle överlappa. Anledningen till den här kompletterande fotograferingen var att skapa bilder som var tänkta att se ut som bilder på ett löpande band i en framtida ostronsorterare skulle kunna se ut.



Figur 2: Ostron av arten *Ostrea* utspridda på en yta som gör det lätt att urskilja varje ostron. Den här typen av bild ansågs i projektet ligga nära hur en framtida implementation av en maskin som sorterar ostron skulle kunna fotografera, där det vita underlaget i framtiden skulle utgöras av något slags rullband. Bild: Jens Wilhelmsson

Bilderna, tillsammans med specifikationen om vilken art som är på vilken bild, användes som träningsunderlag för maskininlärningsalgoritmen. Utifrån bilder likt den i figur 2 klipptes varje enskilt ostron ut med hjälp av identifieringsalgoritmen som beskrivs i nedanstående avsnitt (kap 2.2.1). Totalt användes 3 166 enskilda bilder av ostron för att träna upp och validera algoritmen.

2.2 Genetiska analyser

För att säkerställa vilken art de ostron som betraktades som osäkra var så togs vävnadsprover av mantel och slutarmuskel från de 204 fotograferade individer från den osäkra (US) gruppen. I tillägg togs vävnadsprover från 10 individer var av de säkra grupperna (OE och MG) som kontroller i analysen. Nedan beskrivs i detalj hur de genetiska analyserna genomfördes.

Vävnadsproverna placerades i 96 % etanol och förvarades i -20 °C till analyserna påbörjades. DNA extraherades från varje individ med hjälp av extraktionskitet DNeasy Blood and Tissue (Qiagen). Resultaten från extraktionerna (koncentration av DNA) mättes med en NanoDrop (Thermo Scientific), och några prover testades på en 1 % agarosgel för att utvärdera kvaliteten på det extraherade DNA:t (fragmentlängder).

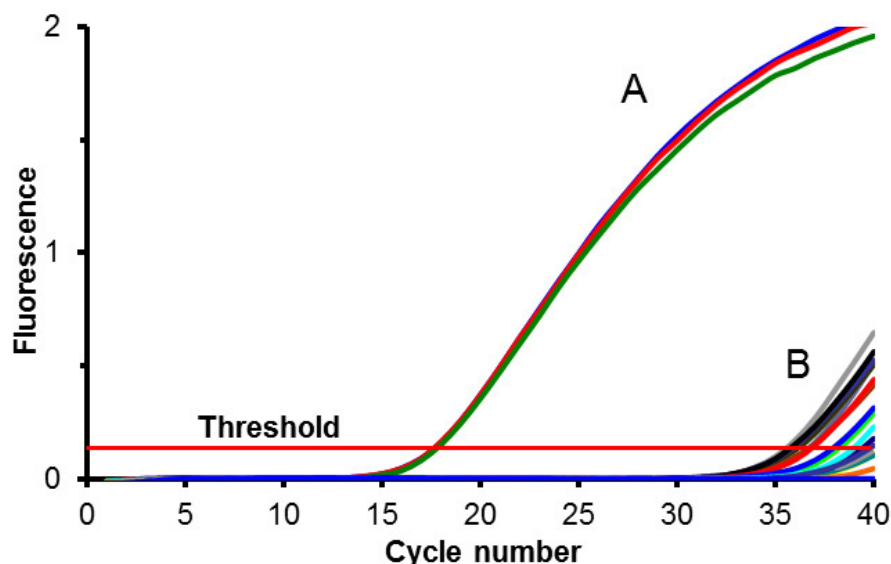
Den ursprungliga planen var att amplifiera och sekvensera delar av COI-genen, med hjälp av universalprimers HCO och LCO (Folmer 1994). Metoden fungerade dock inte optimalt då fler fragment än det önskade DNA-fragmentet amplifierades. Olika tester gjordes för att optimera metoden (t.ex. höjning av temperaturen, beställa nya primers mm.) men utan framgång.

Eftersom sekvenseringen av multipla fragment blir problematisk valdes en alternativ metod baserad på qPCR (quantitative Polymerase Chain Reaction) med artspecifika primers och prober.

Denna metod har tidigare använts för artidentifiering av larver av arterna *O. edulis* och *M. gigas* (Sanchez et al. 2014). Metoden är baserad på amplifiering av DNA i realtid i en PCR-reaktion där fluorescerande prober tillförs vilket gör att PCR-produkten som bildas kan visualiseras genom att mäta fluorescenssignalen i en qPCR-maskin (Figur 3). Två uppsättningar artspecifika primrar samt TaqMan-MGB-prober togs fram för *M. gigas* och *O. edulis* (baserat på sekvenser av 16S rRNA-genen). Proberna märktes med fluorescerande reporterfärger VIC respektive FAM för att kunna skilja mellan arterna. Alla qPCR-reaktioner gjordes enligt protokollet i Sanchez et al (2014), med i en total volym av 20 µl bestående av 10 pl TaqMan Fast Universal PCR MasterMix nr UNG Amperase (2X) (ThermoFisher), 1 µl av varje primer samt prober utifrån koncentrationen som anges i tabell 1, vatten och 2 µl DNA från de extraherade ostronproverna (ca 25 ng µl per L). Varje prov kördes i triplikat på en BioRad q-PCR-maskin i 95 °C under 10 minuter följt av 40 cykler i 95 °C under 15 s, och 60 °C under 1 min. Den fullständiga beskrivningen av metoden finns i Sanchez et al. 2014.

Tabell 1. Primrar och TaqMan-MGB prober sekvenser samt koncentrationer (Sanchez et al. 2014).

TaqMan System	Primer/Probe	Primer Sequence 5' → 3'	Amplicon length (bp)	Primer and Probe concentration (nM)
CGIG/ANG 16S	CGIG/ANG16S_F	GGGCGCCTAGAAAGCAAGT	62	300
	CGIG/ANG16S_R	ATCGGGTCAAATCCGAAAAG		300
	CGIG/ANG16S_P	VIC-AACCTTCTGAATAACTAAC-MGB		200
OEDU 16S	OEDU16S_F	GGCGCCCCACCTAAAAAT	62	900
	OEDU16S_R	AGACCCCGTGCAACTTTTAAAG		900
	OEDU16S_P	FAM-TGAAACTCCTAAACAAGTTG-MGB		225
Positive control	BIV_18S_F	AGCCACACGAGATTGAGCAAT	57	300
	BIV_18S_R	GCGGCCCCGAACATCTA		900
	BIV_18S_P	NED-ACAGGTCTGTGATGCC-MGB		200



Figur 3: Grafisk illustration av hur amplifieringen av DNA från arten av intresse kan se ut i qPCR-maskinen (A) samt amplifiering av andra arter som eventuellt kan finnas i provet (B) (exempel från Sanchez et al. 2014).

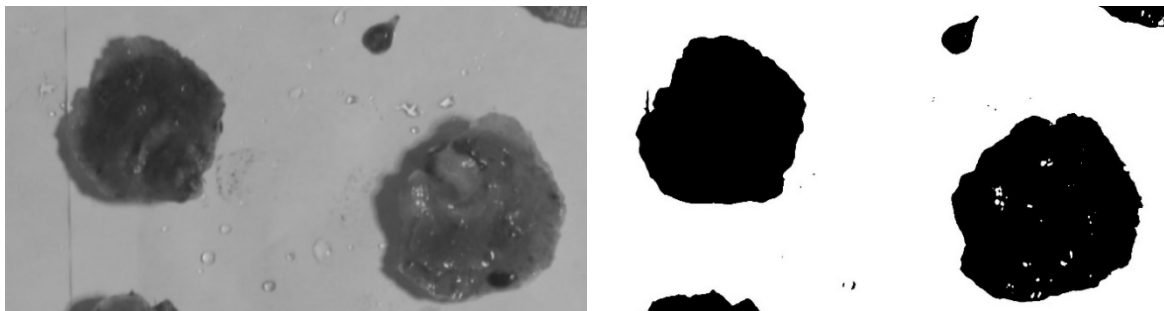
2.3 Bildanalys

För att skapa en programvara som kan klassificera och sortera ostron på löpande band delades arbetet med bildanalysalgoritmen upp i två delar där den ena algoritmen särskiljer varje enskilt ostron i en bild av många ostron (liknande en situation på ett löpande band) och den andra klassificerar de identifierade ostronen som antingen *O. edulis* eller *M. gigas*.

2.3.1 Identifiering av ostron

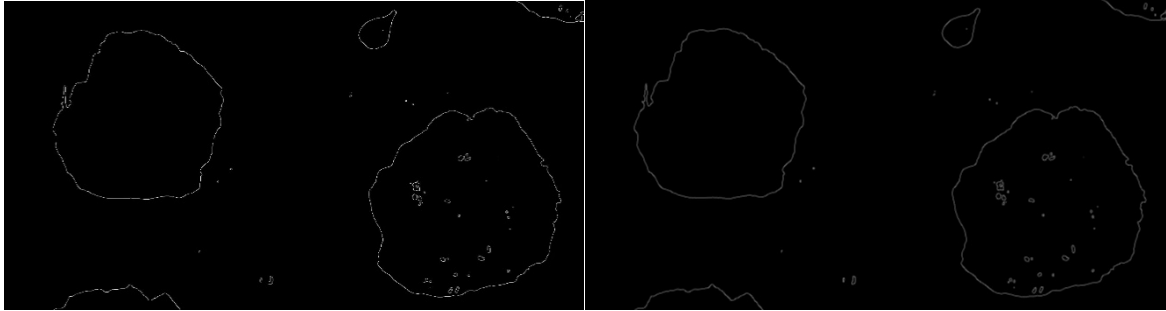
Ett exempel på en bild av många ostron visas i figur 2. De enda villkoren vid fotograferingen av ostron var att inga ostron skulle överlappa och bakgrunden skulle vara enfärgad. Villkoren ansågs i projektet vara rimliga med tanke på en framtida implementering i en fysisk ostronsortare.

Det första steget i att identifiera ostronen var att applicera ett bilateralt utjämningsfilter (Tomasi & Manduchi 1998) vilket är ett filter som jämnar ut bilden men samtidigt bevarar kanter. För att det ska vara enklare att se vad varje steg i processen gör så visas fortsättningsvis endast en liten del av bilden i figur 2. I figur 4 visas resultatet av den bilaterala filtreringen samt gråskalekonvertering till vänster. Till höger i figur 4 ses resultatet av en tröskling av den vänstra bilden. Det innebär att pixlar som har ett värde över ett visst tröskelvärde görs svarta och pixlar med värden under tröskeln görs vita. Ett tröskelvärde hittades med hjälp av Otsus metod (Otsu 1979), en algoritm som används för att ta fram optimala tröskelvärden i bilder.



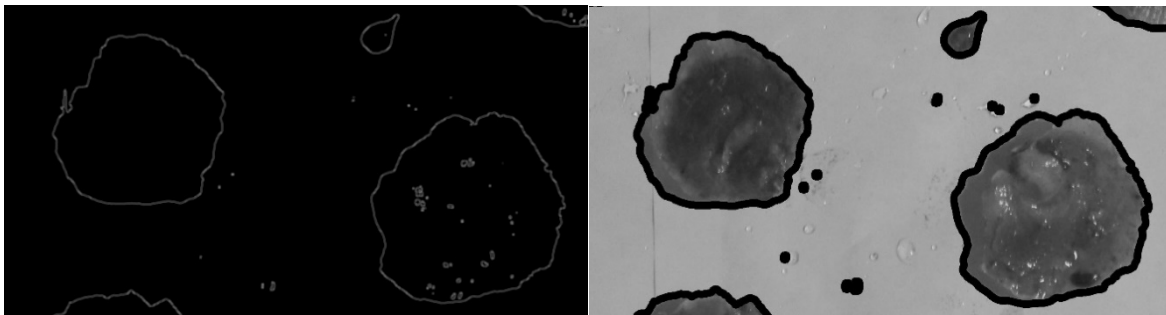
Figur 4: Den vänstra bilden är resultatet av bilateral filtrering, samt gråskalekonvertering. Bilateral filtrering är en utjämningsmetod som bevarar kanter. Den högra bilden är resultatet av tröskling av den vänstra bilden.

Den trösklade bilden användes sedan som underlag till en kantdetekteringsalgoritm, Canny Edge Detector (Canny 1986), vilken bygger på att hitta stora gradienter i bilder (d.v.s. pixlar med värden som skiljer sig mycket från närliggande pixlar). Ytterligare en utjämningsalgoritm som applicerar normalfördelat brus användes sedan för att sluta öppna konturer. Resultaten av båda stegen visas i figur 5.



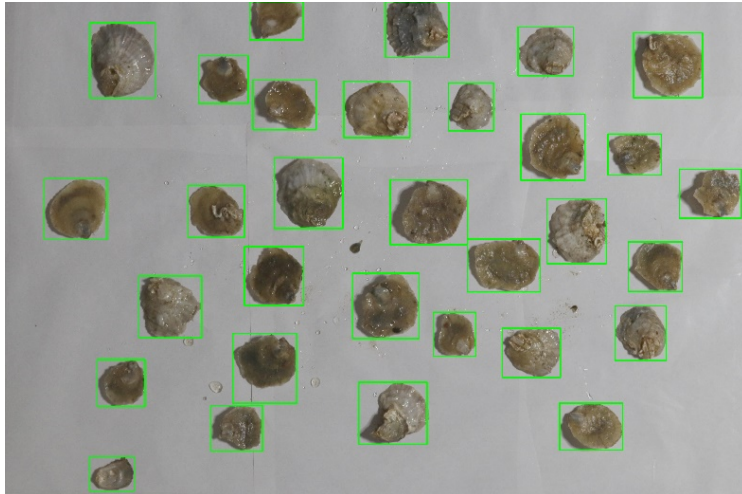
Figur 5: I den vänstra bilden har kantdetektering utförts på den högra bilden i figur 4. Den högra bilden är resultatet av en gaussisk utjämningsalgoritm.

För att ytterligare säkerställa att konturer kring ostron är slutna görs en utvidgning (Haralick & Shapiro 1992) av alla vita pixlar, vilket kan ses till vänster i figur 6. Efter utvidgningen förväntas det att alla konturer kring ostron är slutna, och alla slutna konturer ritas upp i bilden till höger i figur 6.



Figur 6: Utvidgning av alla vita pixlar har skett i den vänstra bilden, och detektion av alla sluta konturer till höger.

Slutligen undersöks alla slutna konturer för att se vilka som är lagom stora för att kunna vara ostron, vilket eliminerar de små prickarna som syns i bilden till höger i figur 6. Utifrån ytterkoordinaterna (upp, ner, höger och vänster) på varje sluten kontur i den högra bilden i figur 6 skapas en rektangel (representerade av gröna rutor), som ses i figur 7. Rektanglarnas koordinater används för att klippa ut varje enskilt ostron.



Figur 7: Resultatet av ostronidentifieringen. De gröna rutorna räknas ut baserat på alla slutna konturer från den högra bilden i figur 6. Varje ostron klipptes ut enligt koordinaterna för de gröna rutorna.

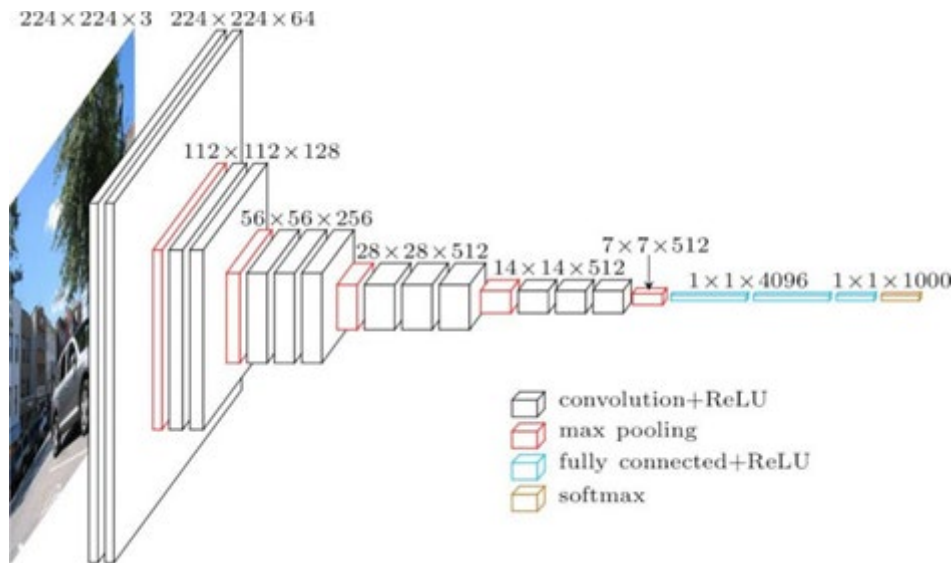
2.3.2 Klassificering av ostron

Nätverket som ostronklassificeringsalgoritmen byggde på var arkitekturen VGG-16 (Simonyan & Zisserman 2014) som visas i figur 7. Det förtränade nätverket tränades av dess skapare inför klassificeringstävlingen ILSVRC2014² där det placerade sig högst upp på topplistan i kategorin klassificering. I klassificeringstävlingen bestod tränings- och utvärderingsdata av bilder från ett dataset vid namn ImageNet³ som består av hundratusentals bilder, fördelade på 1 000 olika kategorier. En bildkategori kan vara allt från till exempel cowboyhatt till zebra. Ingen av kategorierna är ostron, men stora delar av nätverket går ändå att återanvända.

Den sista delen i nätverket (delen längst till höger i figur 8) består av själva klassificeraren medan tidigare lager i arkitekturen tränas att känna igen olika detaljer i bilden. Den typen av struktur gör det möjligt att återanvända majoriteten av ett förtränat VGG-16, där bara de sista lagren byts ut för att passa till att klassificera ostron. I figur 8 går att utläsa att sista lagret har dimensionen 1x1x1 000, där siffran 1 000 syftar på antal bildkategorier i ImageNet. I varianten av VGG16 som använts i det här projektet har det sista lagret dimensionen 1x1x2 eftersom syftet endast var att kategorisera varje ostronbild som antingen *O. edulis* eller *M. gigas*.

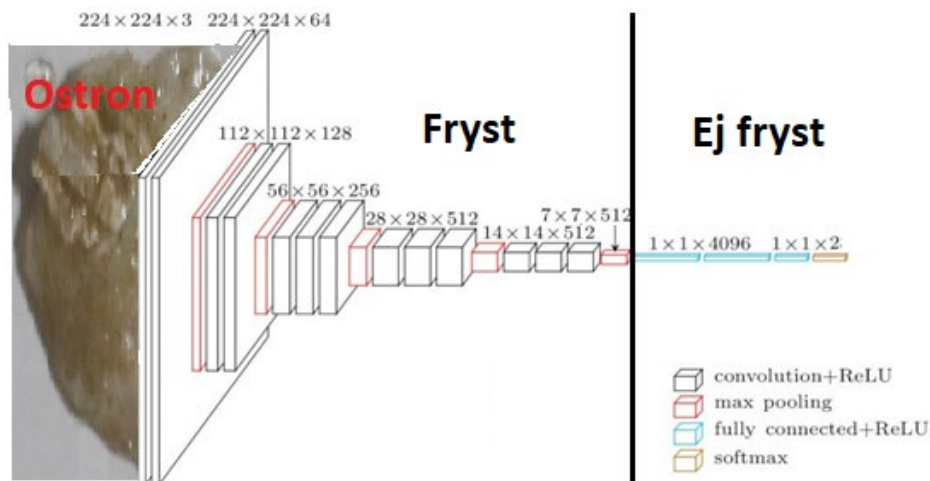
² <http://www.image-net.org/challenges/LSVRC/2014/>

³ <http://www.image-net.org/>



Figur 8: Nätverksarkitekturen VGG16. Notera särskilt det sista lagret som kallas "softmax", den sista siffran 1 000 syftar på antalet kategorier i datasetet ImageNet. Det är i slutet av nätverket som informationen i bilden kopplas till korrekt kategori, medan övriga delar av nätverket specialiserar sig på att utläsa information ur bilden. Det är de avslutande lagren som är kopplade till specifika kategorier som byts ut för att passa till att klassificera ostron (Simonyan & Zisserman 2014).

Träning av ett neuralt nätverk innebär att värden på variabler, som kallas för vikter, uppdateras inkrementellt till dess att nätverket åstadkommit den efterfrågade uppgiften. Att återanvända ett förtränat nätverk innebär med andra ord att värden på vikterna kopieras över till det nya nätverket och att när träningen fortsätter, i detta fall för klassificering av ostron, så fryses större delen av nätverkets tidigare vikter (vilket innebär att vikterna inte kommer att uppdateras, figur 9). Den avslutande delens vikter uppdateras inkrementellt vid träningsförfarandet till dess att ostronklassificeringen uppnår en viss noggrannhet.



Figur 9: Nätverksarkitekturen VGG16, konfigurerad för att tränas till att klassificera ostron. Vikterna i den vänstra delen som är markerad som "frost" hålls konstanta, samtidigt som de vikterna som ej är frysta får uppdateras genom träningsförfarandet (Simonyan & Zisserman 2014).



I slutändan är klassificeringen som nätverket producerar utifrån varje ostronbild ett värde mellan 0 och 1. Ju närmare 1, desto mer säkert är nätverket på att det är *O. edulis*, och ju närmare 0, desto säkrare är nätverket på att det är *M. gigas*.

Följande dataset användes för träning och/eller validering:

Fotografering 1:

1. 80 % av de säkra ostronen. 796 enskilda bilder av ostron.
2. 80 % av de säkra ostronen samt 80 % av de osäkra ostronen (som artbestämdes genom DNA-analys). 1 072 enskilda bilder av ostron.
3. Resterande 20 % av de säkra ostronen samt resterande 20 % av de osäkra ostronen. 222 enskilda bilder av ostron.

Fotografering 2 (många ostron utspridda på en vit yta, alla ostron var säkra i den här fotograferingen):

4. 80 % av ostronen. 1 500 enskilda bilder av ostron.
5. Resterande 20 % av ostronen. 372 enskilda bilder av ostron.

3 Resultat

Resultaten har delats upp så att varje fotografering har varsitt kapitel. Totalt har tre modeller tränats i projektet, varav två har utvärderats på ett gemensamt valideringsdataset.

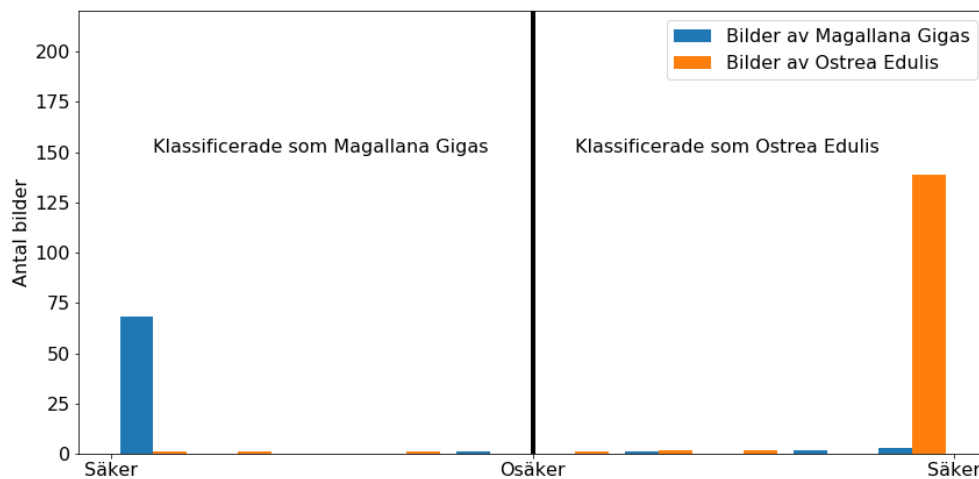
3.1 Fotografering 1

Två likadana klassificeringsmodeller tränades på dataset 1 respektive dataset 2. Anledningen var att se hur inblandningen av svårklassificerade ostron, som klassificerats genom DNA-analys, i träningsdatan skulle påverka prestandan.

3.1.1 Modell tränad på dataset 1

Modellen tränades på dataset 1 som bestod av 796 enskilda bilder av ostron av olika art. Samtliga ostron i datasetet var säkra ostron, vilket innebär att den ursprungliga klassificeringen gjordes utan DNA-analys. Träningen av modellen tog ungefär 40 minuter. Utvärderingen som presenteras nedan är gjord genom att den tränade modellen fått klassificera samtliga bilder i dataset 3, som bestod av både osäkra och säkra ostron.

I figur 10 ses resultatet av klassificeringen i form av ett histogram. De blå staplarna representerar bilder av *M. gigas* medan de orangea staplarna representerar bilder av *O. edulis*. Staplarnas position indikerar resultatet av klassificeringen, längst åt vänster respektive höger innebär säker klassificering av *M. gigas* respektive *O. edulis*. Ju närmare mitten en stapel befinner sig i diagrammet, desto större osäkerhet. I tabell 2 återfinns samma resultat som i histogrammet, men i siffror.

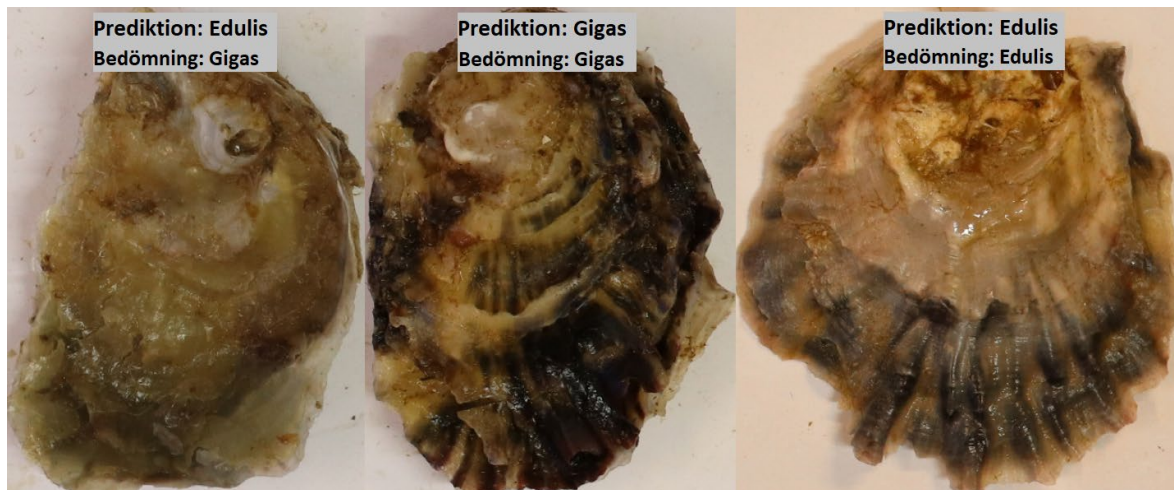


Figur 10: Histogram över resultatet av klassificeringen. Klassificeringen görs i form av en siffra mellan 0 och 1, där 1 betyder *Ostrea edulis* och 0 *Magallana gigas*. Ju närmare 0 och 1, desto säkrare är klassificeringen. Till exempel indikerar de små orangea staplarna till vänster att några bilder av *Ostrea edulis* felaktigt har klassificerats som *Magallana gigas*.

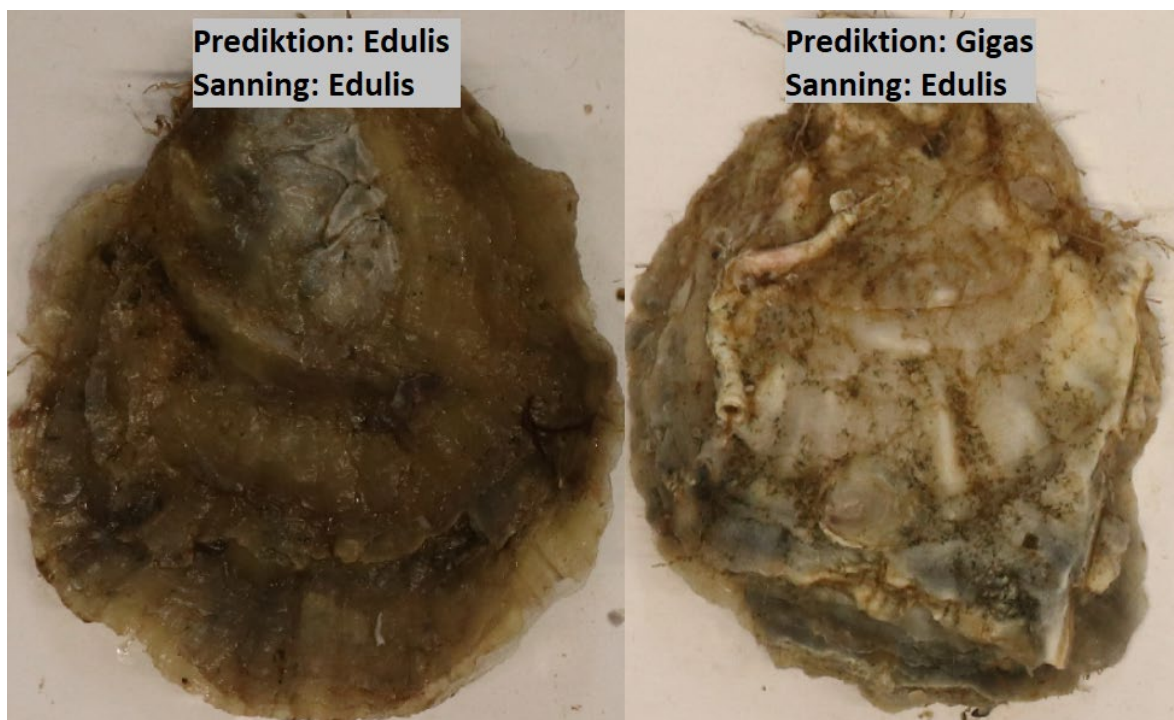
Tabell 2: Resultatet av klassificeringen där modellen som utvärderas har tränats på dataset 1. Bilderna som använts för utvärdering är blandade bilder av säkra ostron och osäkra ostron, där de osäkra artbestämts genom DNA-analys.

	Gigas klassificerade som Gigas	Gigas klassificerade som Edulis	Edulis klassificerade som Edulis	Edulis klassificerade som Gigas	Korrekt klassificerade, alla bilder	Felaktigt klassificerade, alla bilder
Antal	69	6	144	3	213	9
Andel	92,0 %	8,0 %	98,0 %	2,0 %	95,9 %	4,1 %

En majoritet av ostronen (95,9 %) klassificerades korrekt. *M. gigas* klassades i det här försöket oftare fel än vad *O. edulis* gjorde. I figur 11 och 12 visas några exempel på bilder av ostron samt deras prediktioner.



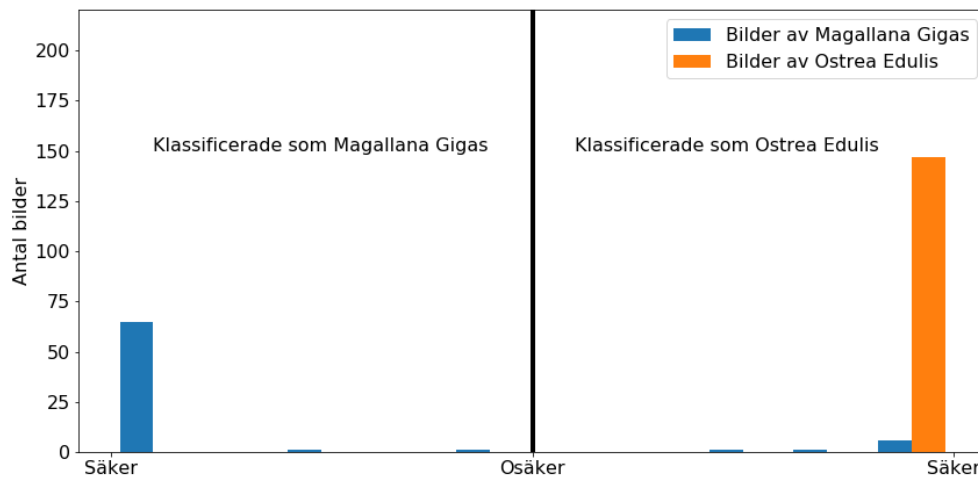
Figur 11: Tre exempel på bilder från valideringsdatasetet (fotografering 1, dataset 3, se kap 2.2.2) i vilket en del av ostronynglena klassats visuellt ("bedömning," se kap 2.1), och den andra delen klassats genom DNA-analys ("sanning"). Samt även vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på endast säkra bilder av ostron (dataset 1, se kap 2.2.2).



Figur 12: Två exempel på bilder från valideringsdatasetet (fotografering 1, dataset 3, se kap 2.2.2) i vilket en del av ostronynglena klassats visuellt ("bedömning," se kap 2.1), och den andra delen klassats genom DNA-analys ("sanning"). Samt även vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på endast säkra bilder av ostron (dataset 1, se kap 2.2.2).

3.1.2 Modell tränad på dataset 2

Modellen i det här avsnittet är av samma typ som i avsnittet ovan, men tränad på dataset 2 istället för dataset 1. Skillnaden mellan dataseten är att det i dataset 2 även ingår bilder av osäkra ostron som artbestämts genom DNA-analys. Resultatet presenteras på samma sätt som i det tidigare avsnittet. I tabell 3 återfinns samma resultat som i figur 9, men i tabellform.

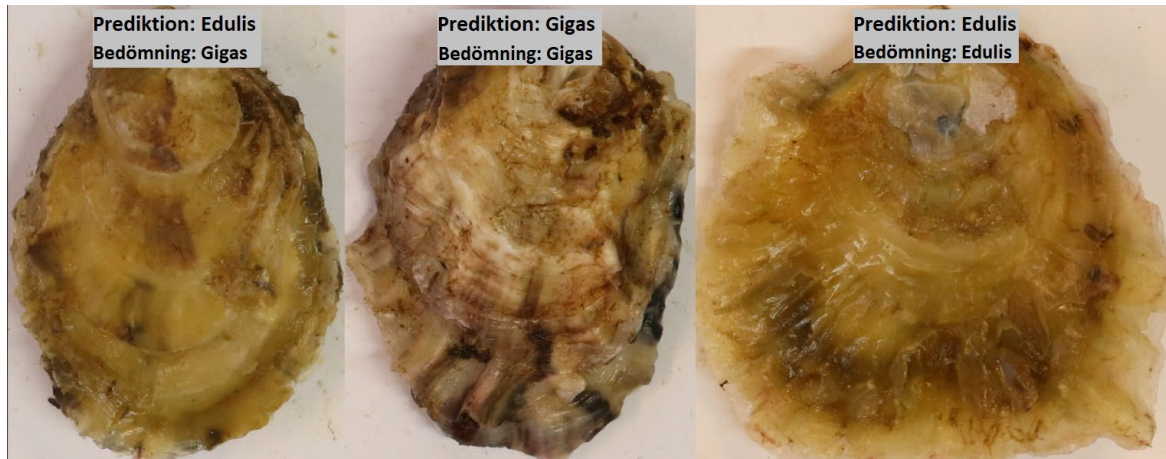


Figur 13: Histogram över resultatet av klassificeringen för modellen tränad på dataset 2. Liksom tidigare görs klassificeringen i form av en siffra mellan 0 och 1, där 1 betyder *Ostrea edulis* och 0 *Magallana gigas*. Ju närmare 0 och 1, desto säkrare är klassificeringen. Till exempel indikerar de små orangea staplarna till vänster att några bilder av *Ostrea edulis* felaktigt har klassificerats som *Magallana gigas*.

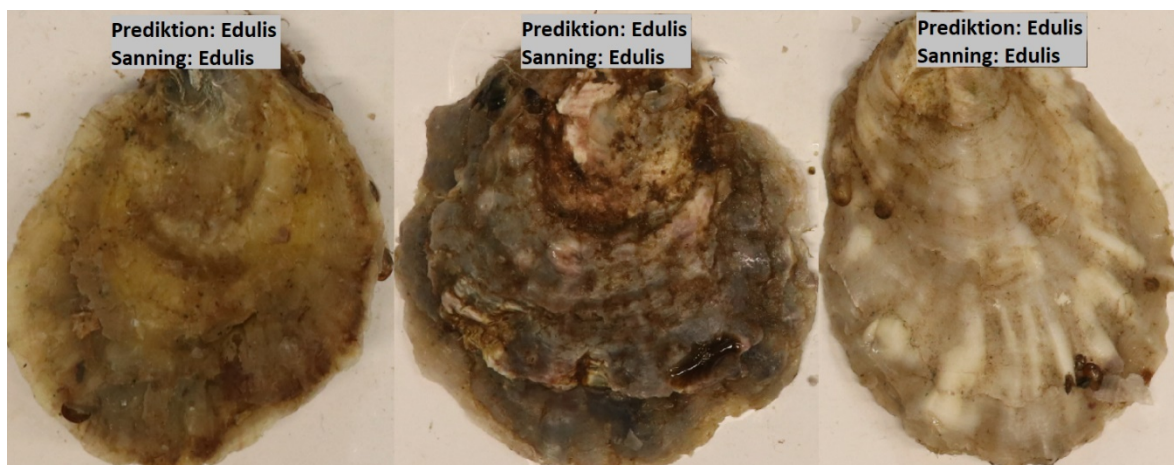
Tabell 3: Resultatet av klassificeringen där modellen som utvärderas har tränats på dataset 2. Bilderna som den utvärderats på nedan är blandade bilder av säkra ostron och osäkra ostron, där de osäkra artbestämts genom DNA-analys.

	Gigas klassificerade som Gigas	Gigas klassificerade som Edulis	Edulis klassificerade som Edulis	Edulis klassificerade som Edulis	Korrekt klassificerade, alla bilder	Felaktigt klassificerade, alla bilder
Antal	67	8	147	0	214	8
Andel	89,3 %	10,7 %	100 %	0%	96,4 %	3,6 %

Ett liknande resultat jämfört med modellen i föregående avsnitt uppnåddes här. Värt att notera är att de enda ostron som klassificerades felaktigt var *M. gigas* som klassificerades som *O. edulis*. I figur 14 och 15 visas några exempel på bilder från valideringsdatasetet.



Figur 14: Tre exempel på bilder från valideringsdatasetet (fotografering 1, dataset 3, se kap 2.2.2) i vilket en del av ostronynglena klassats visuellt ("bedömning," se kap 2.1), och den andra delen klassats genom DNA-analys ("sanning"). Samt även vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på både säkra och osäkra bilder av ostron (dataset 2, se kap 2.2.2).

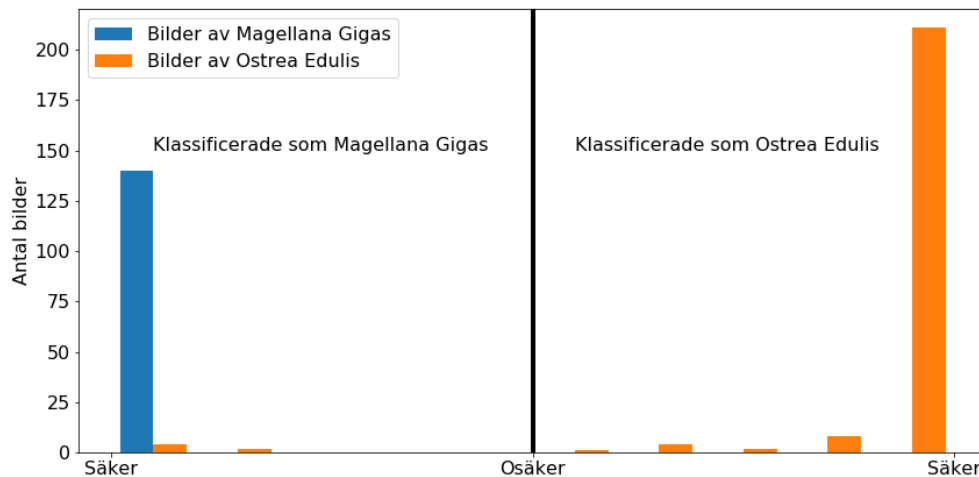


Figur 15: Tre exempel på bilder från valideringsdatasetet (fotografering 1, dataset 3, se kap 2.2.2) i vilket en del av ostronynglena klassats visuellt ("bedömning," se kap 2.1), och den andra delen klassats genom DNA-analys ("sanning"). Samt även vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på både säkra och osäkra bilder av ostron (dataset 2, se kap 2.2.2). I och med att alla osäkra (unsure) var *O. edulis* ("sanning") vilka alla också klassificerades korrekt ("prediktion") av den här modellen, så kan endast bilder av *O. edulis* som klassificerats korrekt visas.

3.2 Fotografering 2

Ostronklassificeringsalgoritmen applicerades först på endast de bilder av ostron där vilken art ostronen tillhörde betraktades som känd. På dessa bilder var ostronen spridda över en vit yta och var inte uppdelade ett och ett på ett rutnät som vid fotografering 1.

I figur 16 ses ett samlat resultat av klassificeringen av alla 372 bilder som användes för validering av den utvecklade klassificeringsalgoritmen. De blå staplarna representerar bilder av *M. gigas* och de orangea staplarna *O. edulis*. Staplarnas position indikerar resultatet av klassificeringen, längst till vänster innebär att bilden med stor säkerhet klassificerats som *M. gigas* och längst till höger innebär att bilden med stor säkerhet klassificerats som *O. edulis*. Ju närmare mitten, desto högre är osäkerheten.



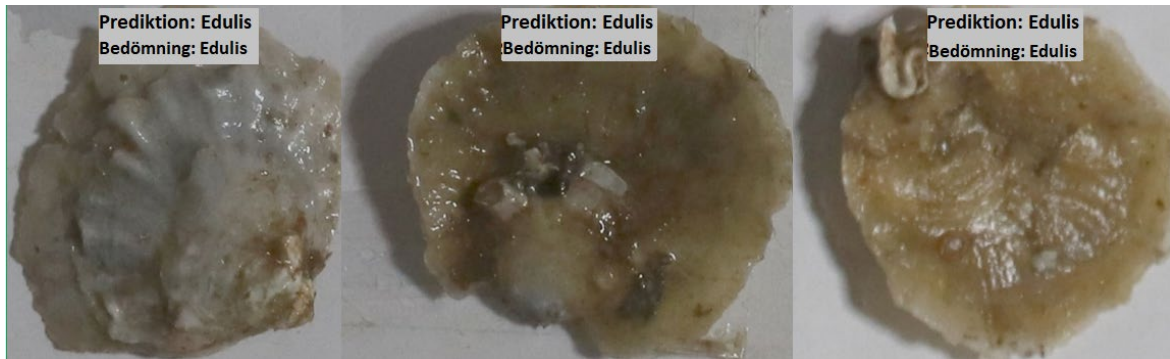
Figur 16: Histogram över resultatet av klassificeringen. Klassificeringen görs i form av en siffra mellan 0 och 1, där 1 betyder *Ostrea edulis* och 0 *Magallana gigas*. Ju närmare 0 och 1, desto säkrare är klassificeringen. De små orangea staplarna till vänster indikerar att några bilder av *O. edulis* felaktigt har klassificerats som *M. gigas*.

En majoritet av ostronen är korrekt klassificerade (tabell 4). Av de som klassificerats felaktigt utgörs samtliga av *O. edulis* som klassificerats som *M. gigas*.

Tabell 4: Resultat från klassificeringen av de ostron som tillhörde de med känd arttillhörighet.

	Gigas klassificerade som Gigas	Gigas klassificerade som Edulis	Edulis klassificerade som Edulis	Edulis klassificerade som Gigas	Korrekt klassificerade, alla bilder	Felaktigt klassificerade, alla bilder
Antal	140	0	226	6	366	6
Andel	100%	0%	97,4%	2,6%	98,4%	1,6%

Ett antal exempelbilder visas i figur 17, 18 och 19. I figur 17 visas tre exempel av *O. edulis* som klassificerats korrekt. I figur 18 visas tre bilder av korrekt klassificerade *M. gigas*. Tre av de ostron som felaktigt klassificerats som *M. gigas* men som är bedömda att vara *O. edulis* visas i figur 19.



Figur 17: Tre ostron som bedömts vara av arten *Ostrea edulis*, korrekt klassificerade. Bilder från valideringsdatasetet (fotografering 2, dataset 5, se kap 2.2.2) i vilket ostronyngel klassats visuellt ("Bedömning", se kap 2.1) samt vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på bilder som extraherats ur gruppbilder av ostron (dataset 4, se kap 2.2.2).



Figur 18: Tre ostron som bedömts vara av arten *Magallana Gigas*, korrekt klassificerade. Bilder från valideringsdatasetet (fotografering 2, dataset 5, se kap 2.2.2) i vilket ostronyngel klassats visuellt ("Bedömning", se kap 2.1) samt vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på bilder som extraherats ur gruppbilder av ostron (dataset 4, se kap 2.2.2).



Figur 19: Tre ostron av arten *Ostrea edulis*, felaktigt klassificerade som *Magallana gigas*. Bilder från valideringsdatasetet (fotografering 2, dataset 5, se kap 2.2.2) i vilket ostronyngel klassats visuellt ("Bedömning", se kap 2.1) samt vad de predikterades som ("prediktion") i utvärderingen av modellen som tränats på bilder som extraherats ur gruppbilder av ostron (dataset 4, se kap 2.2.2).

3.3 Artbestämning genom DNA-analys

Den genetiska qPCR-metoden fungerade mycket bra för att artbestämma ostronen. Resultatet från qPCR visade att alla kontroller (säkra OE och MG) stämde med förväntad art. Av de 168 US prover som analyserades, var majoriteten *O. edulis* (163 individer), medan endast 5 individer var fortsatt osäkra, troligen p.g.a. kontaminering som gör det omöjligt att artbestämma dem. Detta visar att sannolikheten är hög att osäkra ostron med "MG-liknande" egenskaper (ex mörkare färgteckning eller ojämn form) ändå är *O. edulis*. Resultatet visar även att det är relativt enkelt att identifiera och sortera ut MG redan när ostronen är knappt 1 år. Att majoriteten av de osäkra visade sig vara *O. edulis* är inte helt oväntat då ynglen utsattes för ett utbrott av ostronherpesvirus under sommaren 2019. Detta virus orsakar stor dödlighet bland *M. gigas* (dock med minskande dödlighet med ökande storlek av ostron) men påverkar inte *O. edulis*.

4 Diskussion och slutsats

I detta projekt har vi utvärderat möjligheten att använda artificiell intelligens för att utifrån bilder artbestämma ostronyngel som antingen *O. edulis* eller *M. gigas*. Resultatet, sammanfattat i tabell 2, 3 och 4, visar på en mycket stor potential för tekniken. De allra flesta ostronen klassificerades korrekt i samtliga modeller. Till och med modellen som tränades på endast säkra ostron klassificerade 95,9 % av ostronen korrekt på ett valideringsdataset som inkluderade ostron som inte lyckats artbestämmas okulärt utan bestämdes genom DNA-analys. Klassificering baserat både på enskilda och gruppbilder fungerade likvärdigt, vilket indikerar att även om resultaten i denna studie främst är baserade på bilder tagna i en kontrollerad miljö (homogent ljus och ordnad positionering av ostronen för fotografering) så kan ostronyngel troligen klassificeras med hög säkerhet även under mer "ostrukturerade" former e g i samband med storskalig sortering. Detta behöver dock verifieras i samband med vidareutveckling av tekniska lösningar kopplat till ostronsorteringen.

Något förvånande var att klassificeringsframgången knappt skiljde sig mellan modellerna tränade på dataset 1 (endast säkra) respektive dataset 2 (säkra och osäkra inkluderade) när de utvärderades mot samma dataset. Hypotesen var att den modell som även tränats på bilder av osäkra ostron borde prestera bättre, då dessa såg något annorlunda ut än de säkra ostronen. En möjlig anledning till att det blev så är att klassificeraren kan ha tagit fasta på andra visuella egenskaper än vad vi gjorde i vår bedömning, och att dessa möjliggjorde god klassificering även av de ostron som vi bedömt som osäkra. De ostron som visuellt klassificerats som "osäkra", var ofta klassade som troligen *O. edulis* men med egenskaper som påminde om *M. gigas*, vilket gjorde att de inte tydligt kunde placeras bland de "säkra OE". Ett halvår efter att bilderna togs, är många av dessa ostron mer tydligt *O. edulis* utifrån visuell bedömning, vilket också bekräftats med de genetiska analyserna. Detta visar på att *O. edulis* kan ha mer variabelt utseende än vad tidigare bedömts, och att osäkra ostron som inte tydligt kan sorteras ut som *M. gigas*, med stor sannolikhet kan vara *O. edulis*.

För att nå fram till det långsiktiga målet, att konstruera en fysisk ostronsorterare, återstår en del utveckling. Ur ett biologiskt perspektiv bör modellerna utvärderas i förhållande till hur små yngel som programvaran kan identifiera. I detta projekt användes ostron som var drygt 1 år gamla. För framtida utveckling av ett yngelsystem är det önskvärt att sträva efter att yngelsorteringen kan genomföras så tidigt som möjligt (och senast inom ett år) efter att ostronen har bottenfällt (satt sig på kollektorerna, se kap 1.1) eftersom odling av stillahavsosttron inte är tillåtet i dagens läge. Då majoriteten av de yngel som fås på kollektorerna utgörs av *M. gigas* (dock med stora geografiska skillnader) är det av största vikt att yngel som inte kan användas inom produktionen avlägsnas i ett tidigt stadium för att minimera behovet av infrastruktur (odlingskorgar och odlingsriggar) samt arbetsinsats (skötsel) under produktionsfasen. En tidig utgallring av yngel av *M. gigas* (före juni, året efter bottenfällning) minskar också risken för att ynglen av *M. gigas* ska föröka sig då det finns indikationer på att reproduktion kan ske redan under sensommaren året efter att ynglen har bottenfällt om förhållandena är gynnsamma (Faxén 2015).

Steget från att klassificera till att sortera är en uppgift som redan är löst inom många andra fält. Till exempel existerar det sorterare för batterier som också är baserade på bildigenkänning⁴. När det kommer till sortering av blötdjur så finns det liknande system som sorterar musslor, men då oftast

⁴ <https://www.refind.se/optical-battery-sorter-500>

utifrån enklare parametrar såsom färg och form⁵. Ett exempel på en sådan maskin visas i figur 20⁶ (detta är dock endast ett exempel, flera andra företag och modeller finns på marknaden). En bra utgångspunkt för ett fortsättningsprojekt hade därför varit att kombinera tekniker från befintliga sorteringsmaskiner där bildigenkänning används med en bildklassificeringsalgoritm baserad på den som presenterats i detta arbete, dock uppdaterad för mindre ostronyngel.



Figur 20. Exempel på sorteringsmaskin för musslor som använder sig av optisk klassificering⁶.

⁵ <https://www.tomra.com/en/sorting/food/your-produce/seafood/mussels>, <https://www.protec-italy.com/en/sea-food-industry-division/>

⁶ <https://www.tomra.com/en/sorting/food/sorting-equipment/helius>

5 Referenser

- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698.
- FAO 2020. FAO Global Capture Production and FAO Global Aquaculture production. Online statistics database. Tillgänglig: 2020-01-30
- Faxén, A. 2015. Growth and reproduction in one-year old *Crassostrea gigas* - A study in Northern Bohuslän during the summer and autumn of 2014. Masterprojekt, Institutionen för biologi och miljövetenskap, Göteborgs universitet. 39 pp.
- Fish to 2030 - Prospects for Fisheries and Aquaculture (2013). World bank report No. 83177-GLB. The World Bank, Washington. 102 pp.
- Folmer, O., Black, M. Hoeh, W. Lutz, R., Vrijenhoek, R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3(5), 294-299.
- Haralick, R. M., & Shapiro, L. G. (1992). *Computer and robot vision* (Vol. 1, pp. 28-48). Reading: Addison-wesley.
- Hilborn R. Banobi J. Hall S.J. Pucylowski T. Walsworth T.E. 2018. The environmental cost of animal source foods. *Frontiers in ecology and the environment*, 16: 329-335.
- Kotta, J., et al. (2020) Cleaning up seas using blue growth initiatives: Mussel farming for eutrophication control in the Baltic Sea. *Science of the Total Environment* 709: 136144.
- Livsmedelsstrategin 2016. Regeringens proposition 2016/17:104, En livsmedelsstrategi för Sverige – fler jobb och hållbar tillväxt i hela landet.
- OSPAR (2008) Case Reports for the OSPAR List of Threatened and/or Declining Species and Habitats. Biodiversity Series, Pp. 261.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.
- Sanchez, A., Quinteiro J., Rey-Méndez, M., Perez-Martín, R.I. González Sotelo, C. (2014). Identification and quantification of two species of oyster larvae using real-time PCR. *Aquatic living resources*, Vol. 27, Issue 3-4 , pp. 135-145
- SCB 2020. Varuimport från samtliga länder efter varugrupp KN 2,4,6,8-nivå och handelspartner, sekretessrensad, ej bortfallsjusterat. År 1995 – 2018.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Tomasi, C., & Manduchi, R. (1998, January). Bilateral filtering for gray and color images. In *Sixth international conference on computer vision* (IEEE Cat. No. 98CH36271) (pp. 839-846). IEEE.
- UN 2015. *Transforming our world: the 2030 agenda for sustainable development*. 41 pp.
- UN 2019. United Nations, Department of Economic and Social Affairs, Population Division 2019. *World Population Prospects 2019: Ten Key Findings*.
- Vattenbruksstrategin 2012. Svenskt vattenbruk – en grön näring på blå åkrar, strategi 2012–2020. Jordbruksverket. 24 pp.



